3-14-2003

Crystal structure and evolution of a prokaryotic glucoamylase

Alexander E. Aleshin
Iowa State University

Ping-Hua Feng
Iowa State University

Richard B. Honzatko
Iowa State University

See next page for additional authors

Crystal structure and evolution of a prokaryotic glucoamylase

Alexander E. Aleshin[1], Ping-Hua Feng[2], Richard B. Honzatko[1] and Peter J. Reilly[2*]

[1]*Department of Biochemistry, Biophysics, and Molecular Biology and*

[2]*Department of Chemical Engineering, Iowa State University, Ames, Iowa 50011, U. S. A.*

[3]Corresponding author: Department of Chemical Engineering, 2114 Sweeney Hall, Iowa State University, Ames, IA 50011-2230, USA. Telephone: +1-515-294-5968; Fax: +1-515-294-2689. E-mail: reilly@iastate. edu.

Keywords: crystal structure; evolution; glucoamylase; glycoaminoglycan lyase; maltose phosphorylase; *Thermoanaerobacterium thermosaccharolyticum*

Running title: *Prokaryotic Glucoamylase Structure and Evolution*

Abbreviations used: aGA, *Aspergillus niger* glucoamylase; GA, glucoamylase; GAG, glycoaminoglycan; GH, glycoside hydrolase; PEG, polyethylene glycol; tGA, *Thermoanaerobacterium thermosaccharolyticum* glucoamylase

The first crystal structures of a two-domain, prokaryotic glucoamylase were determined to high resolution from the clostridial species *Thermoanaerobacterium thermosaccharolyticum* with and without acarbose. The N-terminal domain has eighteen antiparallel strands arranged in β-sheets of a super-β-sandwich. The C-terminal domain is an $(\alpha/\alpha)_6$ barrel, lacking the peripheral subdomain of eukaryotic glucoamylases. Interdomain contacts are common to all prokaryotic Family GH15 proteins. Domains similar to those of prokaryotic glucoamylases in maltose phosphorylases (Family GH65) and glycoaminoglycan lyases (Family PL8) suggest evolution from a common ancestor. Eukaryotic glucoamylases may have evolved from prokaryotic glucoamylases by the substitution of the N-terminal domain with the peripheral subdomain and by the addition of a starch-binding domain.

Introduction

Glucoamylase (1,4-α-D-glucan glucohydrolase, EC 3.2.1.3, GA) releases β-D-glucose from the nonreducing ends of starch, glycogen, and maltooligosaccharides, cleaving all α-glycosidic bonds between glucosyl residues except that of α,α-trehalose[1]. GAs occur in some prokaryotic and many eukaryotic microorganisms, and may have originated as a polysaccharide exo-hydrolase early in the evolution of glycogen metabolism. Fungal GAs can completely hydrolyze glycogen and raw starch[2]. Industrial processes use GA from *Aspergillus awamori* or *Aspergillus niger* (recently consolidated into *A. niger*) to convert starch to glucose. This enzyme, henceforth aGA, has been the focus of numerous studies in directed mutation and protein engineering[3,4].

All GAs possess $(\alpha/\alpha)_6$-barrel catalytic domains[5,6], known in SCOP nomenclature[7] as 6-helical hairpin toroids. Some yeast GAs have only catalytic domains, but many fungal GAs possess an additional starch-binding domain located either at the N- or C-terminus of the polypeptide chain. C-terminal starch-binding domains have conserved β-sandwich folds of approximately 110 residues[8]. Starch-binding domains of fungal GAs are attached to catalytic domains through flexible, O-glycosylated linkers of up to 40 amino acid residues. *Aspergillus* GAs have the longest linkers, and their starch-binding domains interact weakly with the catalytic domain[9,10].

GAs are part of CAZy Family GH15[11] (http://afmb.cnrs-mrs. fr/~cazy/CAZY). Recently, over 40 open reading frames from bacterial and archaeal genomes have been assigned to this family. All of them have C-terminal catalytic domains, homologous to those of known GAs, but they also possess 200–310-residue domains at their N-termini similar to those of known GAs from prokaryotes. The widespread presence of these domains in GH15 prokaryotic proteins suggests a specific function, but that function is unknown.

A number of inverting carbohydrases, such as maltose phosphorylases/trehalases (Family GH65), exo- and endo-cellulases (Families GH8, GH9, and GH48), and α-1,2-mannosidases (Family GH47), as well as polysaccharide lyases in Families PL5 and PL8, have catalytic domains consisting of toroids with five to seven helical hairpins. Maltose phosphorylase has a catalytic domain with similarities to those of fungal GAs[12], as well as an N-terminal super-β-sandwich domain similar in size to the N-terminal domains of prokaryotic GAs. Structures of prokaryotic GAs therefore might help to establish distant evolutionary relationships between (α/α)-barrel enzymes not evident from sequence data alone.

The biochemical and cellular functions of prokaryotic GAs from thermophilic clostridia have been investigated[13–21], but relatively little is known about them or other prokaryotic GAs. The N-terminal cysteine of *Clostridium* sp. GA is modified during secretion with a fatty acid[17], which may promote GA attachment to the cell wall. A high fraction of the GA produced by these organisms nevertheless can be released to the medium[21].

Here we present the crystal structures of ligand-free and acarbose-bound GA from the thermophilic clostridial species *Thermoanaerobacterium thermosaccharolyticum* (tGA), refined to 2.2 and 2.1 Å resolution, respectively. These are the first tertiary structures reported for any prokaryotic GA. Sequence analysis of Family GH15 proteins are consistent with the evolution of fungal GAs, which have a single catalytic domain, from a two-domain ancestor. Furthermore, bacterial GAs, maltose phosphorylases, and glycoaminoglycan (GAG) lyases may have evolved as well from a common ancestor. A role for the N-terminal β-domain of GAs in substrate exchange is proposed.

Results and Discussion

Overall structure

tGA crystallizes as a dimer with proper twofold symmetry in the space group $P2_12_12_1$ (a = 81.19, b = 102.51, and c = 164.82 Å for ligand-free tGA and a = 80.49, b = 102.93, and c = 164.69 Å for acarbose-ligated tGA). The interface between subunits of the dimer is relatively large (1570 $Å^2$) and contains two regions of direct contact involving eight hydrophobic residues with a total area of 560 $Å^2$ (Figure 1). Hydrophilic residues near the molecular twofold axis are not in direct contact, but interact instead through the water molecules trapped there. The molecular mass of tGA, estimated by gel filtration and sucrose gradient centrifugation, is about 75 kDa, indicating a monomer[16,19–21]. Dimerization of tGA presumably requires elevated protein concentrations and/or stabilizing interactions with other molecules. Indeed, the cell membrane itself could stabilize the dimer. N-terminal cysteines, presumably modified with fatty acids[16], project from the same face of the dimer (Figure 1), with a pair of active sites on the opposite face. In contrast, eukaryotic GAs are monomeric under all conditions examined so far.

tGA consists of three structurally distinct regions (Figure 1): the N-terminal domain (residues 11–252, hereafter the β-domain), the C-terminal domain (residues 295–684, hereafter called the α-domain), and the linker domain (residues 253–294). Electron density is not evident for a 17-residue signal peptide and ten additional N-terminal residues that are either disordered or have been removed by proteolysis.

The α-domain in tGA is an $(\alpha/\alpha)_6$ barrel similar to the catalytic domains of fungal GAs[5,6,22] (Figures 1 and 2). Helical hairpins of the $(\alpha/\alpha)_6$ barrel are arranged into a two-layered toroid with the inner layer containing even-numbered helices (αH2, αH4, etc., where α represents the α-domain) and the outer layer containing odd-numbered helices (αH1, αH3, etc.). A hydrophobic core divides the central region into separate, water-filled voids. Long loops (αL1, αL2, etc.) connect outer with inner helices (αH1 with αH2, αH3 with αH4, etc.), extending one of the voids into a narrow pocket that binds acarbose. In contrast, connecting loops at the other end of

the toroid are relatively short. Loops αL4 and αL5 have intra-loop hydrogen bonds between residues 516–528 and 563–574, respectively, characteristic of antiparallel β-sheets. Loop αL4 participates in two hydrophobic contacts at the subunit interface of the dimer. These contacts are present only in the eubacterial GA subfamily. The α-domain of tGA lacks the small subdomain between α-helices αH10 and αH11 present in GAs from *A. niger* and *Saccharomycopsis fibuligera* (Figure 2).

The 18 antiparallel β-strands of the β-domain divide into two β-sheets (Figures 1 and 2). The first 29 residues (a short β-strand and loop) connect opposite ends of the β-sheets. The rest of the β-domain is classified in the SCOP data base as a super-β-sandwich[7]. Structural homologues revealed by the DALI program[23] include a copper amine oxidase (PDB accession label 1OAC, rmsd 4.5 Å for 189 homologous residues), galactose mutarotase (1L7J, 3.5 Å, 175 residues), methylamine oxidase (1A2V, 3.9 Å, 183 residues), α-mannosidase (1HTY, 3.2 Å, 167 residues), the N-terminal domain of maltose phosphorylase (1H54, 2.9 Å, 166 residues), the C-terminal domains of two GAG lyases (1EGU, 4.1 Å, 163 residues; 1CB8, 4.1 Å, 154 residues), and the fifth domain of β-galactosidase (1BGL, 3.4 Å, 157 residues). The β-super-sandwich folds of copper amine oxidase, galatose mutarotase, and methylamine oxidase are catalytically active by themselves. The β-domain of β-galactosidase participates significantly in intersubunit contacts, but catalysis occurs instead at another domain, an $(\alpha,\beta)_8$-barrel. In contrast, the β-domains of tGA, maltose phosphorylase, and the GAG lyases interact extensively with $(\alpha/\alpha)_6$ barrels within their own subunits, but their functions are otherwise uncertain.

The contact area between domains within a subunit of the dimer is 2240 $Å^2$, approximately 18% of the total surface area of the separate domains. The α-domain contributes 55 residues to the interface from helices αH1 and αH2 and loops αL1 and αL6. The β-domain contributes 49 residues from strands βS2–βS4 and associated loops. The second of two helices of the linker (residues 269–285) interacts with strand βS18 and helix αH1.

GAs from *A. niger* and *S. fibuligera* have no β-domain and yet are five- to tenfold more active than tGA[3,21,24]. So what benefit do prokaryotic GAs derive from their β-domains? Glycosyl-

ation and the aforementioned peripheral subdomain stabilize the catalytic domain of eukaryotic GAs[25,26]. The glycan attached to Asn395 of aGA loop αL6[5] covers the surface corresponding to the interdomain interface of tGA (Figure 2). This glycosylation site is conserved among fungal GAs[6]. The mutation of Asn395 in aGA significantly reduces secretion and thermostability, but does not affect catalysis[27]. Similarly, recombinant *S. fibuligera* GA expressed in *Escherichia coli* does not fold properly[24]. *In vitro* refolding results in active protein, but with reduced thermostability[24]. Hence, the β-domain of tGA, as an alternative to glycosylation and the peripheral subdomain, may direct the correct fold and/or confer thermostability to the catalytic domain.

Evolutionary relationships among Family GH15 proteins

Sequences and open reading frames from all 32 Family GH15 prokaryotic proteins surveyed have weakly conserved N-terminal segments (the initial 200–310 residues) followed by homologous α-domains (the complete sequence alignment is in the supplemental material). Two isozymes from *Arthrobacter globiformis* have additional 330-residue extensions of unknown fold and function to their C-termini. Family GH15 sequences fall into four major subfamilies (Figure 3): six bacterial GAs (Subfamily 1), including tGA; seven archaeal and one bacterial GA (Subfamily 2); 13 hypothetical proteins from bacteria, archaea, and fungi (Subfamily 3); and fungal GAs, many of which were classified earlier[6] but are not shown in Figure 3 (Subfamily 4). Two highly divergent open reading frames from *Clostridium acetobutylicum* and *Sulfolobus tokodaii* (b12 and a13 of Figure 3) may be part of additional subfamilies.

The N-terminal domains in Subfamilies 1 and 2, which include biochemically characterized GAs, align with gaps not likely to disrupt the super β-sandwich fold. The N-terminal segments from Subfamily 3 have poor homology with β-domains of other subfamilies; however, strands βS2–βS4, βS18, and the last helix in the linker domain have consensus sequences in Subfamilies 1–3 (Figure 3). Residues from the α-domain in contact with the β- and linker domains are conserved also. Hence, the intra-subunit interface between α-, β-, and linker domains may be common to Subfamilies 1–3. Contrary to known starch- and cellulose-binding domains, the N-

terminal domains of prokaryotic GAs do not have conserved surface residues. Hence, Family GH15 β-domains are unlikely to have common substrate-binding sites.

All proteins in Subfamilies 1, 2, and 4 are probably GAs, as they have conserved catalytic and active-site residues. The hypothetical enzymes of Subfamily 3, on the other hand, may be carbohydrases, but possibly not GAs. They have "catalytic" glutamates, but several other residues important for GA activity and substrate binding are not present: a hydrophobic residue replaces Glu439 in tGA (Glu180 in aGA), Gly or Ser replace Glu434 (Asp176 in aGA), and Glu or His replace Gln637 (Gln401 in aGA). Moreover, several archaeal and fungal organisms from *Schizosaccharomyces*, *Sulfolobus*, and *Thermoplasma* have proteins from Subfamily 3 together with GAs from Subfamilies 2 or 4, while organisms unable to hydrolyze starch, such as *Mycobacterium tuberculosis* and *Streptomyces coelicolor*, nevertheless have proteins in Subfamily 3.

Differences in sequence and function suggest divergence early in evolution between Subfamilies 1, 2, and 4 on the one hand and Subfamily 3, b12, and a13, on the other. Moreover, all kingdoms of microorganisms have proteins in Subfamily 3, whereas each of the other subfamilies has proteins primarily from a single kingdom. These observations support the assignment of Subfamily 3 as an outgroup for GAs (Figures 3 and 5d). Even though low bootstrap values (Materials and Methods) indicate substantial uncertainty in the evolutionary order of GA subfamilies, the evolution of fungal GAs from a two-domain antecedent is inescapable, given the choice of outgroup. This conclusion is unexpected, as evolution usually goes from the simple to the complex; however, as presented below a two-domain fold probably existed long before the appearance of Family GH15. Several factors may have driven the evolution of single-domain fungal GAs (Subfamily 4) from a two-domain ancestor: The β-domain may confer extreme thermostability, a property of diminished value to organisms that exist in relatively gentle environments. Hence, a conserved glycosylation site and a conserved α-helical subdomain (Figure 2) may have provided adequate compensation for the loss of the β-domain. In addition, the β-domain may have encumbered the translation and secretion mechanism of fungi to no advantage.

Structural homology among GH15, GH65, and PL8 families

tGA, maltose phosphorylase (PDB entry 1H54), and three GAG lyases (PDB entries 1CB8, 1EGU, and 1F1S) consist of topologically homologous α- and β-domains with conserved inter-domain contacts (Figure 4). Maltose phosphorylase and tGA are exo-acting enzymes with similar active sites and catalytic mechanisms[12]. Their N-terminal β-domains connect to α-domains through two-helix linkers. GAG lyases are endo-acting and break glycosyl bonds by β-elimination[28]. Their β-domains are C-terminal, and their catalytic domains are distorted (α/α) barrels (Figures 4 and 5). In hyaluronate lyase, the helix homologous to αH12 of tGA is displaced from the inner layer of the toroid[29], whereas in chondroitin lyase the corresponding helix is absent, yielding a horseshoe-like α-domain[30]. Lyase (α/α) barrels have cleft-shaped active sites characteristic of endo-acting carbohydrases, rather than the blind-hole pockets of exo-hydrolases.

Despite the differences above, three β-strands of GAG lyases and maltose phsophorylase are homologous to strands βS2–βS4 of tGA, and interact with an α-helix and active-site loops that include residues of functional significance: Glu636 in tGA, His671 in maltose phosphorylase, and His399 in hyaluronate lyase (Figure 4). The conserved location of the β-domains relative to the active site suggests a similar role in these enzymes. Indeed, the binding of phosphate to the active site of maltose phosphorylase causes considerable movement of the β-domain together with the aforementioned loops[12]. Similar changes were proposed in response to substrate binding to hyaluronate lyase[28]. As shown below, the α- and β-domains of tGA also have different relative positions in ligand-free and ligated structures.

The structural similarities between prokaryotic GAs and maltose phosphorylases imply a common two-domain antecedent in evolution; however, an evolutionary relationship between GAG lyases and GAs is less obvious, because their β-domains have different topological relationships to their (α,α) barrels (Figure 5a,b). The topology of GAG lyases may have arisen through recombination of an ancestral tGA gene (or *vice-versa*), as depicted in Figure 5c, which did not affect the interaction between the domains. Single α-domain lyases from Family PL5[11]

may have evolved from PL8 lyases by loss of the β-domain, just as eukaryotic GAs may have evolved from prokaryotic GAs (Figure 5d).

Interactions with acarbose

Up to six glucosyl residues at the nonreducing end of substrates influence tGA kinetics[20]. The tGA/acarbose complex reveals the first and second glucosyl subsites deep within the pocket formed by loops αL1–αL6. The first subsite holds the nonreducing end of acarbose and the catalytic water without additional empty space. The active site constricts around the first glucosidic linkage and then widens into the second subsite, which is large enough to accommodate an α-(1→6)-branched polysaccharide (Figure 6). GAs in general bind and hydrolyze branched substrates, allowing complete digestion of starch. Amino acid residues that hydrogen-bind acarbose (Figure 6, Table 1) are on the side of the active site distal to the β-domain. Hydrophobic residues, four of which are tryptophans, cover the hydrophobic faces of the first two acarbose residues. Two other subsites are located on the enzyme surface and provide hydrophobic interactions to only one face of the third and fourth rings of acarbose. The last residue of acarbose binds to the other GA molecule in the dimer through a mediating sulfate ion (Table 1).

Even though the sequence identity in the α-domains of tGA and aGA is only 17%, their active sites are virtually identical. The rmsd of all atoms in the 12 residues of subsites 1 and 2 in tGA and aGA is 0.38 Å, while the rmsd of their corresponding Cα atoms is 0.25 Å. The Cα atom of Trp390 (Trp120 in aGA) has a different relative position, being displaced by 3.3 Å. Nevertheless, the side chains of Trp390 (tGA) and Trp120 (aGA) have similar interactions with the third residue of acarbose.

The conformation of acarbose in the tGA structure is intermediate between the two acarbose conformations observed in the aGA structure[31]. Modeling suggests that in the absence of the sulfate ion, acarbose may bind to loop αL3, similar to its binding mode in aGA at pH 4[31]. In that structure, the hydrophobic faces of the third and fourth acarbose residues are covered by loop αL3, while the 4-OH group of the fourth residue hydrogen-bonds to the backbone carbonyl 180

(backbone carbonyl 439 in tGA).

Catalytic mechanism

GAs hydrolyze the glycosidic bond of the nonreducing end of polysaccharides by a general acid-base mechanism[32]. Glu438 of tGA (Glu179 in aGA) and Glu636 (Glu400 in aGA) are the putative catalytic acid and base, respectively (Figure 6). Glu438 probably protonates the oxygen atom of the glycosidic bond, resulting in bond cleavage and forming a glucopyranosyl cation. The water molecule trapped in the first subsite attacks the anomeric carbon on the side opposite that of the leaving group, producing β-glucose. The catalytic base assists hydrolysis by abstracting a proton from the catalytic water. This mechanism, generally accepted for inverting enzymes, has been confirmed in aGA by directed mutations of the catalytic residues[33,34].

Acarbose is a potent GA inhibitor ($K_i \sim 10^{-12}$ M)[4] because it mimics some of the properties of the transition state. The imino nitrogen atom of acarbose is probably protonated and positively charged when it binds to GA. The nonreducing end of acarbose has a double bond between atoms C-5 and C-7 (the latter replaces the endocyclic oxygen atom), and resembles the half-chair conformation of the glucopyranosyl cation. An electrostatic calculation by DelPhi[35] infers a strong negative potential in the active site that would stabilize the transition state and contribute to the tight binding of acarbose. The rigid active sites observed so far in all GA structures, including tGA, are consistent with a substrate distortion mechanism[31]. Tyr337, Trp341, and the catalytic water in tGA have close contacts with the C-6 and C-7 atoms of the nonreducing end of acarbose (Figure 6, Table 1). A nonreducing end in the $^4C_1$ conformation makes even tighter contacts with the corresponding aGA residues[31].

Conformational differences between ligand-free and acarbose-bound tGA

The bottle-shaped active site of tGA, observed in its acarbose complex, sterically blocks the substrate from the first subsite; however, the active sites of all GAs, ligand-free or otherwise, have nearly identical conformations. In the absence of ligands, seven ordered water molecules

occupy the first subsite, of which six are expelled by acarbose. The active-site opening in ligand-free tGA is too small for water molecules to exit as the substrate enters the active site. Perhaps the active sites of GAs open transiently during ligand exchange.

Observed differences between the ligand-free and ligated crystal structures of tGA support the hypothesis above. Evidently, the α-domain rotates by 2° relative to the β-domain about an axis located at the N-terminus of the α-domain (Figure 7). In addition, loops αL1–αL6 shift toward each other upon ligation, further closing the active site. In the ligand-free structure, residues 331–336 of loop αL1 interact weakly with the β-domain and have thermal parameters averaging 60 $Å^2$. In the ligated structure the same residues interact with a different region of the β-domain and their average thermal parameter falls to 40 $Å^2$. Trp390, its side chain disordered in the ligand-free structure, shifts along with loop αL2 by 0.9 Å, and provides a hydrophobic contact with the third residue of acarbose (Figure 7). The side chains of Tyr337 and Trp341 also shift toward each other, closing the entrance to the first subsite by 0.6–1.4 Å. More secure interactions between the α- and β-domains then correlate with a more closed active site. Hence, the loss of interactions between the β-domain and loop αL1 could open the active site sufficiently to permit ligand exchange.

The absence of the β-domain in eukaryotic GAs may reflect a different mechanism of ligand exchange. Structures of *A. niger* and *S. fibuligera* GAs have extensive networks of water-filled cavities in their catalytic domains. Approximately 40 water molecules are under loops αL1–αL5 in aGA[36]. Water clusters extend laterally from the first subsite to the enzyme surface. The water-filled cavities in aGA might facilitate transient movements in the loops during substrate exchange. *S. fibuligera* GA is evolutionarily distant from aGA (30% sequence identity in the catalytic domain) and yet many of the water molecules in their voids occupy common relative positions (27 out of 40 water molecules under loops αL2–αL4 have an rmsd of less than 0.3 Å). In contrast, tGA has fewer than 20 water molecules under loops αL1–αL6, and the only large water-filled cavity (eight water molecules) is located under loop αL3.

Conclusions

Similarities in structure and/or function between prokaryotic GAs, maltose phosphorylase, and the GAG lyases suggest evolution from a common, two-domain ancestor. Although the active sites lie entirely (as in GA and maltose phosphorylase) or mostly (as in GAG lyases) within their α-domains, the β-domain could stabilize a closed, productive conformation of the active site. GAG lyase and maltose phosphorylase crystal structures also suggest large movements of their domains over the course of a catalytic cycle. In prokaryotic GAs the active site may open transiently to facilitate ligand exchange.

A major event in the evolution of eukaryotic GAs may have been the loss of the β-domain and a concomitant change in the mechanism of ligand exchange. The selective pressure that favored the loss of the β-domain in eukaryotic GAs may have originated from differences in secretion mechanisms between eukaryotes and prokaryotes, as well as extensive glycosylation of secreted eukaryotic proteins. Most prokaryotic GAs come from thermophilic organisms; a harsh environment may select for the stabilizing influence of the β-domain. Enhanced thermostability of tGA may be due to the β-domain, as well as to the reduced number of water molecules trapped under the loops forming the active site.

Materials and Methods

Preparation of protein and crystals

tGA was produced and purified from *T. thermosaccharolyticum* ATCC 7956 (DSM 571) as described previously[21]. tGA crystals were grown at 24°C using the hanging-drop/vapor-diffusion method, where 2 μL of a 14 g/L tGA solution in 20 mM Tris-HCl buffer, pH 7.5, was mixed with the same volume of a reservoir solution containing 14–16% PEG (polyethylene glycol) 3350, 100 mM Tris-HCl buffer, pH 8.0, and 200 mM Li_2SO_4. Thin plates, 0.2 x 0.2 x 0.02 mm, grew within 1–2 weeks and were then transferred to a storage solution containing 16% PEG, 100 mM Li_2SO_4, and 50 mM sodium/potassium phosphate buffer, pH 7.0. Low sequence homology

between tGA and fungal GAs and poor tGA production by *T. thermosaccharolyticum* precluded structure determinations by MR and selenomethionine MAD. MAD phasing from a single heavy-atom derivative (12-h soak of the native crystal with 0.1 mM K_2PtBr_4) ultimately provided a structure. A solution complex of tGA/acarbose did not crystallize. Native crystals soaked overnight in 10 mM acarbose cracked severely, but the cracks later disappeared, leaving the diffraction properties of the crystal undiminished and unit cell parameters unchanged (Results and Discussion). Native, acarbose-soaked, and Pt-derivatized crystals were flash-frozen after short exposure to the original soaking/storage solution mixed with 25% (v/v) glycerol.

Data collection and phase determination

MAD data from the Pt-derivative crystal and high-resolution data from the native crystal were collected at 100 K on beamline 17-ID (IMCA-CAT) of the Advanced Photon Source, Argonne, Ill. Three wavelengths were selected for MAD data on the basis of an X-ray fluorescence spectrum collected from the crystal. Data for the tGA-acarbose complex were collected at 100 K on beamline 19-BM (SBC-CAT) of the same synchrotron. All data sets were processed using program HKL[37]. Data collection statistics are in Table 2.

Seven heavy atoms were located by SOLVE[38], which then was used to calculate phases. Initial phases had a figure-of-merit of 0.47 to a resolution of 2.8 Å. A self-rotation function, calculated with program MolRep of CCP4[39], revealed noncrystallographic twofold symmetry. The orientation and position of the noncrystallographic symmetry axis was found with program FindNCS of CCP4, using heavy-atom coordinates. Solvent flattening (estimated solvent content of 35%), noncrystallographic averaging, and histogram fitting was performed with program DM[39]. The mask for noncrystallographic averaging was built with programs NCSMASK[39] and O[40]. After density modification, the figure-of-merit was 0.83 to a resolution of 2.8 Å.

Refinement

A model was built into an electron density map based on DM-modified phases, using pro-

grams O[40] and Xfit of Xtalview[41]. The initial model was refined with program CNS[42] using the experimental phases and noncrystallographic restraints. Calculated phases were combined with experimental phases, producing an electron density map of excellent quality, which was used to complete model building. The model was transferred to the unit cell of the high-resolution native data set by the CNS rigid-body refinement and then refined against data to 2.1 Å resolution by slow-cooling simulated annealing with a starting temperature of 5000°C. All subsequent refinement cycles were done using simulated annealing at a constant temperature of 500°C, followed by conjugated gradient energy minimization of coordinates and individual thermal parameters. Restraints on noncrystallographic symmetry were used during all refinement cycles, gradually reducing their weights to 30 and 8 for coordinates and thermal parameters, respectively, by the end of the refinement. Water molecules were added automatically using CNS and a 3-σ cut-off for peaks in $F_{obs} - F_{calc}$ maps. The upper limit on hydrogen bond distances to neighboring atoms was 3.4 Å and the thermal parameter limit was 70 Å^2. Water molecules were added until the R_{free} factor stopped improving. The refined model of the native structure served as the initial model for the acarbose-ligated structure. Its refinement and water-building procedures were as with the native structure. Topology and parameter files, as well as the initial model for acarbose, came from the HIC-Up database[43].

Model quality and refinement statistics

Refinement statistics are presented in Table 2. Model quality was validated with CNS and Procheck[44]. Categories of stereochemistry are as good or better than those of structures of comparable resolution. No residues fall into disallowed regions of the Ramachandran plot, and only two residues of each monomer are in generously allowed regions. Each monomer contains four *cis*-prolines validated by omit maps.

Coordinates have been deposited in the Protein Data Bank (accession codes 1LF6 for tGA and 1LF9 for tGA complexed with acarbose).

Phylogenetic and structural comparisons

Protein sequences of Figure 4 were aligned with programs Clustal X[45] and Pileup of GCG[46], followed by manual corrections using the crystal structures of tGA and aGA. The phylogenetic analysis of aligned α- and β-domains employed the Fitch-Margoliash method and parsimony as implemented in the Phylip package[47]. Distances for the former method were corrected using the Dayhoff PAM matrix[48]. Sequence insertions longer than three amino acids were excluded from the aligned sequences. Both methods gave consistent results. Bootstrap analysis gave good reliability of phylogeny inference within subfamilies (bootstrap values generally above 95% and in all cases above 50%), but low reliability (bootstrap values below 50%) for nodes representing the divergence of Subfamilies 1–4 from each other.

Structural homologues of the tGA β-domain were retrieved from the DALI server and superimposed on tGA using Lsqkabch of CCP4[39]. The same program was used to analyze conformational differences between ligand-free and acarbose-ligated tGA.

Acknowledgements

The authors thank the staff of IMCA and SBC of the Advanced Photon Source, Argonne National Laboratory, for providing beam time and assistance in data collection. Grants from the Plant Sciences Institute of Iowa State University and the National Institutes of Health (NS10546) supported Alexander Aleshin, and grants from the U. S. Department of Energy through the Consortium for Plant Biotechnology Research, Inc. as well as from Genencor International, Inc. supported Ping-Hua Feng.

References

1. Meagher, M. M. & Reilly, P. J. (1989). Kinetics of the hydrolysis of di- and trisaccharides by *Aspergillus niger* glucoamylases I and II. *Biotechnol. Bioeng.* **34**, 689–693.
2. Manjunath, P., Shenoy, B. C. & Raghavendra Rao, M. R. (1983). Fungal glucoamylases. *J.*

Appl. Biochem. **5**, 235–260.

3. Reilly, P. J. (1999). Protein engineering of glucoamylase to improve industrial performance–A review. *Starch/Stärke* **51**, 269–274.
4. Sauer, J., Sigurskjold, B. W., Christensen, U., Frandsen, T. P., Mirgorodskaya, E., Harrison, M., Roepstorff, P. & Svensson, B. (2000). Glucoamylase: structure/function relationships, and protein engineering. *Biochim. Biophys. Acta* **1543**, 275–293.
5. Aleshin, A., Golubev, A., Firsov, L. M. & Honzatko, R. B. (1992). Crystal structure of glucoamylase from *Aspergillus awamori* var. *X100* to 2.2-Å resolution. *J. Biol. Chem.* **267**, 19291–19298.
6. Coutinho, P. M. & Reilly, P. J. (1997). Glucoamylase structural, functional, and evolutionary relationships. *Proteins: Struct. Funct. Genet*. **29**, 334–347.
7. Lo Conte, L., Ailey, B., Hubbard, T. J., Brenner, S. E., Murzin, A. G. & Chothia, C. (2000). SCOP: a structural classification of proteins database. *Nucleic Acids Res*. **28**, 257–259.
8. Sorimachi, K., Jacks, A. J., Le Gal-Coëffet, M. -F., Williamson, G., Archer, D. B. & Williamson, M. P. (1996). Solution structure of the granular starch binding domain of glucoamylase from *Aspergillus niger* by nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **259**, 970–987.
9. Williamson, G., Belshaw, N. J., Noel, T. R., Ring, S. G. & Williamson, M. P. (1992). *O*-Glycosylation and stability. Unfolding of glucoamylase induced by heat and guanidine hydrochloride. *Eur. J. Biochem*. **207**, 661–670.
10. Khan, S. M. A., Reilly, P. J. & Ford, C. (2000). Thermal and molecular characterization of *Aspergillus awamori* glucoamylase catalytic and starch-binding domains. *Starch/Stärke* **52**, 385–397.
11. Coutinho, P. M. & Henrissat, B. (1999). Carbohydrate-active enzymes: An integrated database approach. In *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G., Henrissat, B. & Svensson, B., eds.), pp. 3-12. The Royal Society of Chemistry, Cambridge.

12. Egloff, M.-P., Uppenberg, J., Haalck, L. & van Tilbeurgh, H. (2001). Crystal structure of maltose phosphorylase from *Lactobacillus brevis*: unexpected evolutionary relationship with glucoamylases. *Structure* **9**, 689–697.
13. Hyun, H. H. & Zeikus, J. G. (1985). General biochemical characterization of thermostable pullulanase and glucoamylase from *Clostridium thermohydrosulfuricum. Appl. Environ. Microbiol*. **49**, 1168–1173.
14. Hyun, H. H. & Zeikus, J. G. (1985). Regulation and genetic enhancement of glucoamylase and pullulanase production in *Clostridium thermohydrosulfuricum. J. Bacteriol.* **164**, 1146–1152.
15. Ohnishi, H., Sakai, H. & Ohta, T. (1991). Purification and some properties of glucoamylase from *Clostridium* sp. *G0005*. *Agric. Biol. Chem*. **55**, 1901–1902.
16. Specka, U., Mayer, F. & Antranikian, G. (1991). Purification and properties of a thermo-active glucoamylase from *Clostridium thermosaccharolyticum*. *Appl. Environ. Microbiol.* **57**, 2317–2323.
17. Ohnishi, H., Kitamura, H., Minowa, T., Sakai, H. & Ohta, T. (1992). Molecular cloning of a glucoamylase gene from a thermophilic *Clostridium* and kinetics of the cloned enzyme. *Eur. J. Biochem.* **207**, 413–418.
18. Specka, U. & Mayer, F. (1993). Cellular location, activity states, and macromolecular organization of glucoamylase in *Clostridium thermosaccharolyticum*. *Arch. Microbiol*. **160**, 284–287.
19. Ducki, A., Grundmann, O., Konermann, L., Mayer, F. & Hoppert, M. (1998). Glucoamylase from *Thermoanaerobacterium thermosaccharolyticum*: Sequence studies and analysis of the macromolecular architecture of the enzyme. *J. Gen. Appl. Microbiol*. **44**, 327–335.
20. Ganghofner, D., Kellermann, J., Staudenbauer, W. L. & Bronnenmeier, K. (1998). Purification and properties of an amylopullulanase, a glucoamylase, and an α-glucosidase in the amylolytic enzyme system of *Thermoanaerobacterium thermosaccharolyticum*. *Biosci. Biotechnol. Biochem*. **62**, 302–308.

21. Feng, P.-H., Berensmeier, S., Buchholz, K. & Reilly, P. J. (2002). Production, purification, and characterization of *Thermoanaerobacterium thermosaccharolyticum* glucoamylase. *Starch/Stärke* **54**, 328–337.
22. Ševcik, J., Solovicová, A., Hostinová, E., Gašperík, J., Wilson, K. S. & Dauter, Z. (1998). Structure of glucoamylase from *Saccharomycopsis fibuligera* at 1.7 Å resolution. *Acta Crystallogr.* **D54**, 854–866.
23. Holm, L. & Sander, C. (1996). Alignment of three-dimensional protein structures: network server for database searching. *Methods Enzymol.* **266**, 653–662.
24. Solovicová, A., Gašperík, J. & Hostinová, E. (1996). High-yield production of *Saccharomycopsis fibuligera* glucoamylase in *Escherichia coli*, refolding, and comparison of the nonglycosylated and glycosylated enzyme forms. *Biochem. Biophys. Res. Commun.* **224**, 790–795.
25. Evans, R., Ford, C., Sierks, M., Nikolov, Z. & Svensson, B. (1990). Activity and thermal stability of genetically truncated forms of *Aspergillus* glucoamylase. *Gene* **91**, 131–134.
26. Libby, C. B., Cornett, C. A. G., Reilly, P. J. & Ford, C. (1994). Effect of amino acid deletions in the *O*-glycosylated region of *Aspergullus awamori* glucoamylase. *Protein Eng.* **7**, 1109–1114.
27. Chen, H.-M., Ford, C. & Reilly, P. J. (1994). Substitution of asparagine in *Aspergillus awamori* glucoamylase by site-directed mutagenesis to eliminate N-glycosylation and inactivation by deamidation. *Biochem. J.* **301**, 275–281.
28. Jedrzejas, M. J. (2000). Structural and functional comparison of polysaccharide-degrading enzymes. *Crit. Rev. Biochem. Mol. Biol.* **35**, 221–251.
29. Li, S., Kelly, S. J., Lamani, E., Ferraroni, M. & Jedrzejas, M. J. (2000). Structural basis of hyaluronan degradation by *Streptococcus pneumoniae* hyaluronate lyase. *EMBO J.* **19**, 1228–1240.
30. Féthière, J., Eggimann, B. & Cygler, M. (1999). Crystal structure of chondroitin AC lyase, a representative of a family of glycosaminoglycan degrading enzymes. *J. Mol. Biol.* **288**, 635–647, doi:10.1006/jmbi.1999.2698.

31. Aleshin, A. E., Stoffer, B., Firsov, L. M., Svensson, B. & Honzatko, R. B. (1996). Crystallographic complexes of glucoamylase with maltooligosaccharide analogs: relationship of stereochemical distortions at the nonreducing end to the catalytic mechanism. *Biochemistry* **35**, 8319–8328.
32. Sinnott, M. L. (1990). Catalytic mechanisms of enzymatic glycosyl transfer. *Chem. Rev.* **90**, 1171–1202.
33. Sierks, M. R., Ford, C., Reilly, P. J. & Svensson, B. (1990). Catalytic mechanism of fungal glucoamylase as defined by mutagenesis of Asp176, Glu179, and Glu180 in the enzyme from *Aspergillus awamori. Protein Eng.* **3**, 193–198.
34. Frandsen, T. P., Dupont, C., Lehmbeck, J., Stoffer, B., Sierks, M. R., Honzatko, R. B. & Svensson, B. (1994). Site-directed mutagenesis of the catalytic base glutamic acid 400 in glucoamylase from *Aspergillus niger* and of tyrosine 48 and glutamine 401, both hydrogen-bonded to the γ-carboxylate group of glutamic acid 400. *Biochemistry* **33**, 13808–13816.
35. Nicholls, A. & Honig, B. (1991). A rapid finite difference algorithm, utilizing successive overrelaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.* **12**, 435–445.
36. Aleshin, A. E., Hoffman, C., Firsov, L. M. & Honzatko, R. B. (1994). Refined crystal structures of glucoamylase from *Aspergillus awamori* var. *X100*. *J. Mol. Biol.* **238**, 575–591, doi:10.1006/jmbi.1994.1316.
37. Otwinowsiki, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. **276**, 307–326.
38. Terwilliger, T. C. & J. Berendzen (1999). Automated structure solution for MIR and MAD. *Acta Crystallogr.* **D55**, 849–861.
39. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for crystallography. *Acta Crystallogr.* **D50**, 760–763.
40. Jones, T. A., Zou, J. -Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr*. **A47**, 110–119.

41. McRee, D. E. (1992). A visual protein crystallographic software system for X11/Xview. *J. Mol. Graph*. **10**, 44–46.
42. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr*. **D54**, 905–921.
43. Kleywegt, G. J. & Jones, T. A. (1998). Databases in protein crystallography. *Acta Crystallogr*. **D54**, 1119–1131.
44. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
45. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The Clustal X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. **24**, 4876–4862.
46. Womble, D. D. (2000). GCG: the Wisconsin Package of sequence analysis programs. *Methods Mol. Biol.* **132**, 3–22.
47. Felsenstein, J. (1989). PHYLIP–phylogeny inference package (Version 3. 2). *Cladistics* **5**, 164–166.
48. Dayhoff, M. O., Eck, R. V. & Park, C. M. (1972). A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure*, Vol. 5 (Dayhoff, M. O., ed.), pp. 89–99, National Biomedical Research Foundation, Washington.

Table 1. Selected contacts between acarbose and tGA

Atom in acarbose	Atom in tGA[a]	Distance (Å)[b]
O2A	NH2 Arg575 (Arg305)	2.8/2.7
O3A	NE Arg343 (Arg54)	3.1/3.2
	Carbonyl Arg436 (Leu177)	2.7/2.7
O4A	NH2 Arg343 (Arg54)	2.7/2.7
	NE Arg343 (Arg54)	3.2/3.2
	OD2 Asp 344 (Asp55)	2.8/3.0
C6A	CD2 Trp341 (Trp52)	3.5/3.6
O6A	OD1 Asp344 (Asp55)	2.6/2.5
	Catalytic water	3.1/3.0
C7A	OH Tyr337 (Tyr48)	3.3/3.3
O2B	OE2 Glu439 (Glu180)	2.7/2.6
O3B	NH2 Arg575 (Arg305)	2.8/2.5
	Carbonyl Trp437 (Trp178)	2.6/2.6
N4B	OE1 Glu438 (Glu179)	2.5/2.6
O6C	NE2 Gln380	2.9/2.9
	Carbonyl Glu438 (Glu179)	2.9/3.0
O5D	O2 SO_4 (—)	3.5/2.9
	O3 SO_4 (—)	2.9/3.4

O6D	O3 SO_4 (—)	—/3.3
	O4 SO_4 (—)	—/3.1

[a]Corresponding residue from aGA follows in parentheses.

[b]Listed are distances for each subunit of the dimer.

Table 2. Data collection and refinement statistics

Data set	Native	Acarbose complex	Pt Derivative[a]
Wavelength (Å)	1.01	1.00	1.07229
Resolution (Å)	30-2.1	30-2.2	30-2.75
Observations	446,637	268,721	276,500
Unique reflections	76,199	65,095	33,664
Completeness (%)[b]	90 (82)	92 (83)	93 (88)
R_{sym} (%)[b,c]	3.5 (7)	11 (40)	5.7 (27)
<*I*/sigma>[d]	38.5	13.8	18.8
Refinement[d]*:*			
R-factor/R_{free}-factor (%)[e]	19/23	19/23	
Number of protein atoms	10666	10666	
Number of solvent atoms	911	674	
Rms deviations:			
Bond length (Å)	0.006	0.006	
Bond angles (°)	1.2	1.2	
Average B-parameters (Å^2)*:*			
Wilson plot	22	32	
Protein	25	36	
Acarbose		36	

[a]Only statistics for the last collected data set of the Pt derivative is shown, as the other two sets collected at wavelengths of 1.07072 and 1.05072 Å have similar or better parameters.

[b]The values in parentheses refer to the highest resolution shell.

[c]$R_{sim} = \Sigma \mid I_j - \langle I_j \rangle \mid / \Sigma \langle I_j \rangle$, where I_j is the intensity of an observation of reflection j and $\langle I_j \rangle$ is the average intensity of reflection.

[d]All measured reflections were used for refinement.

[e]R-factor = $\Sigma \mid\mid F_{obs} \mid - \mid F_{calc} \mid\mid / \mid F_{obs} \mid$. R_{free}-factor was calculated using 7% of reflections omitted from refinement.

Figure Legends

Figure 1. Overview of the tGA structure. (a) Perpendicular views of tGA dimer complexed with acarbose (van der Waals spheres) viewed down the dimer twofold axis (left) and rotated by 90° about the vertical (right). Possible interactions with the cell membrane may promote dimerization. Regions of direct contact between hydrophobic side chains are in boxes. (b) Schematic representing folding topology and the relationship of sequence to secondary structure. β-Strands that interact with the α-domain are in black. That region of the β-domain within the dotted outline is conserved in other enzymes. Numbers identify start and end residues of secondary elements. The tGA model is based on amino acid residues deduced from a nucleotide sequence[19] (GenBank accession code GI3243238). The crystal structure, however, contains an additional Phe residue at position 125 and five residues after position 678. These residues are present in the homologous (95% identical) GA from *Clostridium* sp. G0005[17]. The reported deletions in tGA may be due to sequencing errors.

Figure 2. Structures of tGA and the aGA α-domain. (a) Stereo view of the tGA β-domain. (b) Structural correspondence between the β-domain and linker region of tGA (left) and the N-glycan of aGA (right). Tan and red loops are on the catalytic side of the α-domain. Loops αL1 and αL6 of tGA that interact with the β-domain and linker are red, as are loops that interact with a corresponding N-glycosylation site of aGA. Catalytic residues are blue. Acarbose in the tGA active site is gray. Conserved N-glycosylation of fungal GAs (aGA-specific O-glycosylation sites are not shown for clarity) is in black. The conserved subdomain of fungal GAs consisting of helix αH10' (previously labeled helix 11[5]) and a β-strand hairpin is in dark green.

Figure 3. Sequence-based relationships between GAs and GA-like proteins. (a) Sequence conservation in secondary elements that participate in the domain interface within subunits. Strands βS2 and loops αL1 and αL6 are omitted to conserve space. Asterisks above the alignment designate residues in contacts between domains. The letters *s* and *h* below the alignment indicate

residues belonging to the corresponding structural elements in tGA. Symbols to the left correspond to the sequences defined above. All glycine and proline residues are highlighted in light brown and yellow, respectively. Residues exhibiting 75% conservation are color-coded as follows: hydrophobic/hydrophobic-polar, light/dark blue; acidic, purple; basic, dark brown; and small-polar, green. (b) A phylogenetic tree where subfamilies labeled 1–4 are as follows: (1) bacterial GAs, (2) archaeal GAs, (3) GA-like hypothetical proteins, and (4) fungal GAs. The filled triangle represents a possible root of the tree. Species designations are (a) archaeal, (b) bacterial, and (f) fungal. Underlined designations are GAs for which crystallographic structures are available. The unrooted phylogenetic tree was built from α-domain sequences. Branch lengths correspond to evolutionary distances. Open reading frames of a13 and b12 may represent two additional subfamilies. Slightly divergent isozymes are represented by a single branch. The sequences used in alignment and phylogenetic analysis are *Sulfolobus solfataricus* (GeneBank GI numbers 13814174, 13815775, 13816070, 13816085) (a1, a5, a9, a10), *Sulfolobus tokodaii* (15621834, 15623127, 15921380, 15921215, 15921362) (a2, a6, a8, a11, a13), *Thermoplasma volcanium* (14324643) (a3), *Thermoplasma acidophilum* (160814739, 16082505) (a4, a12), *Methanococcus jannaschii* (1592211) (a7), *Thermoactinomyces vulgaris* (8777462) (b1), *Agrobacterium tumefaciens* (15890159) (b2), *Mesorhizobium loti* (13473562) (b3), *A. globiformis* (6939848) (b4), *T. thermosaccharolyticum* (3243238) (b5), *Streptomyces coelicolor* (10803154, 4757095, 5738493, 6117873) (b6–b9), *Corynebacterium glutamicum* (12542968) (b10), *Mycobacterium tuberculosis* (1655657) (b11), *Clostridium acetobutylicum* (15025851) (b12), *A. niger* (67393) (f1), *Aspergillus oryzae* (543806, 83675) (f2, f3), *Neurospora crassa* (5979236) (f4), *S. fibuligera* (626994) (f5), *Schizosaccharomyces pombe* (3006163, 1184016) (f6, f8), and *Arxula adeninivorans* (600385) (f7). Sequences for *A. globiformis*, *Clostridium sp. G0005*, and *M. tuberculosis* (1655477, 216417, and 13882189, respectively) are not shown because they are very closely related to b4, b5, and b11, respectively.

Figure 4. Structurally homologous carbohydrases consisting of α- and β-domains. (a) tGA. (b)

Maltose phosphorylase (PDB entry 1H54). (c) Hyalorunate lyase (PDB entry 1EGU). α- and β-domains and linker regions are tan, blue and green, respectively. Additional C-terminal domains of maltose phosphorylase and hyaluronate lyase are brown. Conserved interfacial β-strands and α-helices are dark blue and red, respectively. Glu438 and Glu636, the catalytic acid and base of tGA; Glu487 and His671, the putative catalytic acid and phosphoryl binding residue of maltose phosphorylase; and His399, a proton acceptor of hyalorunate lyase, are black.

Figure 5. Topological and evolutionary divergence of structurally related carbohydrases. (a) Schematic of tGA and maltose phosphorylase. (b) Schematic of hyaluronate lyase. Additional C-terminal domains of maltose phosphorylase and hyaluronate lyase are not shown for clarity. β-Domains are boxes with interfacial β-strands shown as arrows. (c) Possible mechanism of divergent evolution of GAs, maltose phosphorylases (MP), and GAG lyases (HL). The gene segments corresponding to the β-domain and helices αH5–αH12 of tGA and maltose phosphorylase switch places to produce the topology of hyaluronate lyase. Recombination does not affect the domain interface, but changes the relative positions of two interfacial loops and a helix (αL1, αL6, αH1 vs. αL4, αL5, αH9). (d) Possible evolution of an antecedent, two-domain protein into the present-day carbohydrase families. Circles: α-domains of GAs and maltose phosphorylase; cut circles: α-domains of lyases; squares: β-domains.

Figure 6. Interactions of acarbose with the tGA active site. Proposed location of a α-1,6-linked glucosyl branch residue is in thin lines. Hydrogen bonds are dotted lines.

Figure 7. Conformational differences between ligand-free (dark blue and green) and acarbose-complexed (light blue and tan) tGA. (a) Superposition of the β-domains reveals a 2° rigid-body rotation of the α-domain. Black dots represent areas of largest conformational change. (b) Superposition of the α-domains shows conformational changes in the active site. Acarbose is in black and red. A black circle represents the catalytic water. Red dotted lines are hydrogen bonds.

Supplemental Materials

Multiple alignment of Family 15 sequences. The sequences used in the alignment and phylogenetic analysis are *Sulfolobus solfataricus* (GeneBank GI numbers 13814174, 13815775, 13816070, 13816085) (a1, a5, a9, a10), *Sulfolobus tokodaii* (15621834, 15623127, 15921380, 15921215, 15921362) (a2, a6, a8, a11, a13), *Thermoplasma volcanium* (14324643) (a3), *Thermoplasma acidophilum* (160814739, 16082505) (a4, a12), *Methanococcus jannaschii* (1592211) (a7), *Thermoactinomyces vulgaris* (8777462) (b1), *Agrobacterium tumefaciens* (15890159) (b2), *Mesorhizobium loti* (13473562) (b3), *A. globiformis* (6939848) (b4), *T. thermosaccharolyticum* (3243238) (b5), *Streptomyces coelicolor* (10803154, 4757095, 5738493, 6117873) (b6–b9), *Corynebacterium glutamicum* (12542968) (b10), *Mycobacterium tuberculosis* (1655657) (b11), *Clostridium acetobutylicum* (15025851) (b12), *A. niger* (67393) (f1), *Aspergillus oryzae* (543806, 83675) (f2, f3), *Saccharomycopsis fibuligera* (626994) (f5). Sequences for *A. globiformis*, *Clostridium sp. G0005*, and *M. tuberculosis* (1655477, 216417, and 13882189, respectively) are not shown because they are very closely related to b4, b5, and b11, respectively. Asterisks above the alignment designate catalytic acid and base. The letters *s* and *h* below the alignment indicate residues belonging to the corresponding structural elements in f1 and b5. Symbols to the left correspond to the sequences defined above. All glycine and proline residues are highlighted in light brown and yellow, respectively. Residues exhibiting 75% conservation are color-coded as follows: hydrophobic/hydrophobic-polar, light/dark blue; acidic, purple; basic, dark brown; and small-polar, green. C-terminal domains of b2, b3, and b4 and starch-binding domains of f1, f2, and f3 have been omitted.

(a)

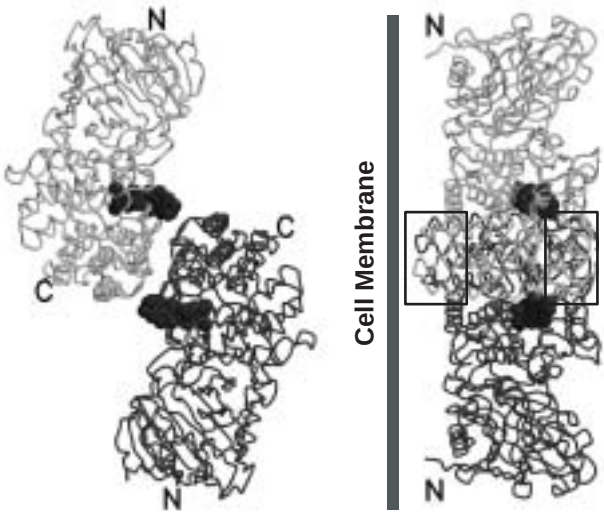

N
C
C
N
Cell Membrane
N
N

Figure 1

(b)

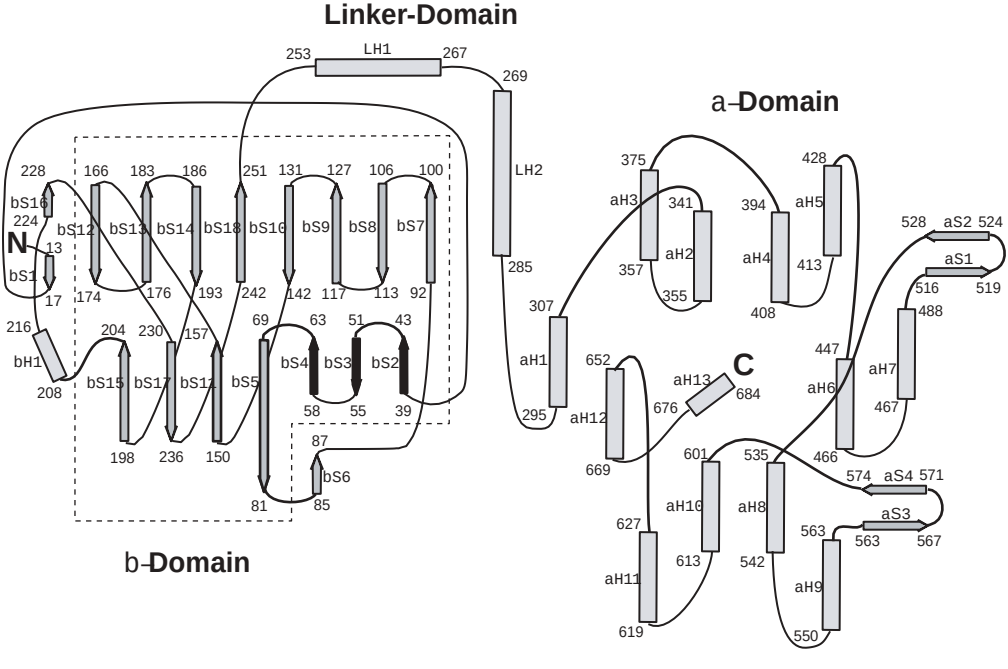
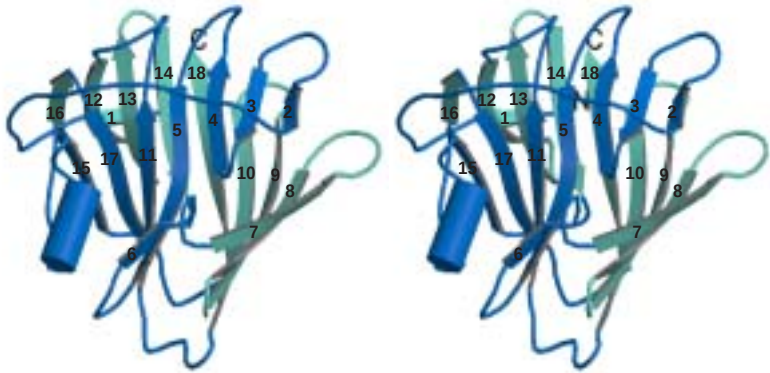

Linker-Domain
LH1
253
267
269
LH2
285
a–Domain
b–Domain
N
C
bS16
bS12
bS13
bS14
bS18
bS16
bS9
bS8
bS7
bS1
bH1
bS15
bS17
bS11
bS5
bS4
bS3
bS2
bS6
aH1
aH2
aH3
aH4
aH5
aH6
aH7
aH8
aH9
aH10
aH11
aH12
aH13
aS1
aS2
aS3
aS4
228
224
13
17
216
208
166
183
186
251
131
127
106
100
174
176
193
242
142
117
113
92
204
230
157
69
63
51
43
58
55
39
198
236
150
87
81
85
375
341
357
355
394
428
413
408
307
295
652
669
676
684
447
466
467
488
516
519
528
524
574
571
563
567
601
627
613
619
535
542
550

Figure 2

(a)


C
14
18
16
12
13
1
5
4
3
2
15
17
11
10
9
8
7
6
C
14
18
16
12
13
1
5
4
3
2
15
17
11
10
9
8
7
6

(b)

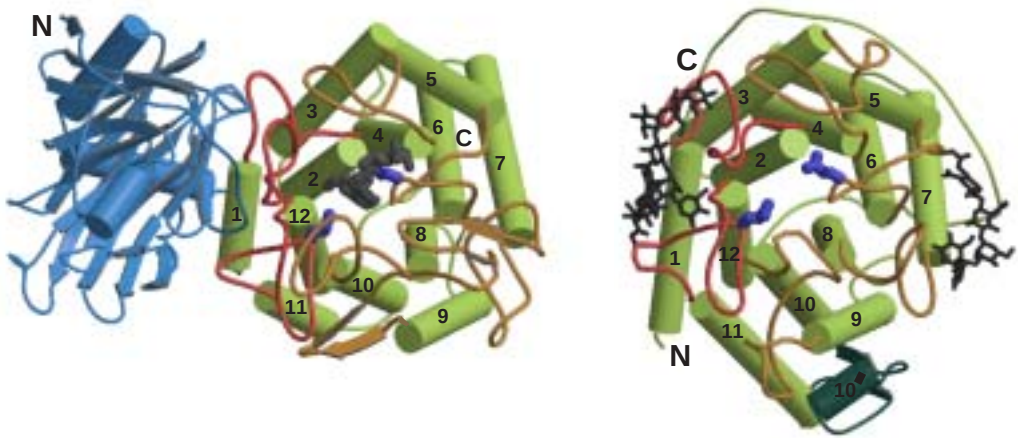

N
5
3
6
C
4
7
2
1
12
8
10
11
9
C
3
5
4
2
6
7
1
12
8
10
9
11
N
10

Figure 3

(a)

```
     *        * *    *   *          * * * *       *  *   *  *** **       *   **  **  *
a8  AIIQ-GTSIVWFPVPKF---DS  30   GFLVANYSTD 164   EKPFEKTIK-YWESFE 196   YEDLYKSSVFVLLG 215
a9  ALID-LSSIVWFPVPKF---DS  34   GYLIANYASD 170   ERSFENTIN-YWKSID 202   FNDLYKASIYTMLG 222
a10 LIK-NGSV-EWLPLPRF---DS  32   GYIYLLYSKD 172   E-ALNKAIR-YWRRRL 203   YVDAYYRSLLVLLG 225
a11 LES-QGSI-DWFPVPRF---DS  47   GYLYLLYSKD 185   E-AFSKLIA-LSEKEL 216   FKDIYYRSISVLLG 241
 b8 LVCRDGTV-DWLCLPRF---DS  48   VAFTISWEP- 198   S-LV--ATEDFWRDWV 224   YREAVVRSLITLKA 246
 b9 LVSADGSV-DWLCLPRF---DS  36   VAFVLTWHP- 183   MS-LEHSVE-DWREWA 209   HRDAVVRSLITLKA 231
b10 LLSNMGSL-DWLCLPRF---DS  49   VEWVLTWAP- 202   RS-LESTLS-FWASWV 228   YDAEVRRSMLVLRA 250
b11 LISPAGSV-EWLCVPRP---DS  47   VFVALSWTK- 218   ADKMWQTTE-CWRQWI 243   WRAYLQRSALTLKG 266
 b7 LVDRRGAI-DWMCVPRF---DS  32   FGFSLVWAP- 185   AERIDDTVE-AWRSWE 214   HRDLVLLSSRVLKG 237
a12 LVAMNGFI-DWGCLPNF---NS  51   SWVIALYGI- 202   L-RLQETTD-YWRKWA 228   YHSMVMRSALALKV 250
 f8 MISLDGSV-EMMCWPNF---DS  50   QEITFIFRQ- 218   VDKLEDSTKRYWRAWI 248   YREFVQRNALTLKL 270
 b6 VLDPRGRI-VWLCAPRW--HDD  53   VLELSAGRE- 202   ADALWRATEQEWRRAV 223   RDARHAYA----VL 242
 a1 NFDEKGRIVDIYY-P-YIGMEN  33   LKLYYVIVA- 214   VETSFTLTYMFWRNWL 248   IKRVYDVSLFVIRN 275
 a2 NIDEFGRLTDLYF-P-YVGMEN  33   DKAYFVLAF- 211   VETTFTSNYMFWQNWI 245   VEKLYKISLLVIKN 269
 a3 AFDEDYRIVDFYF-SKYAS-EN  57   FTVYYFIISG 243   IRKLVKRTTNYWELWS 276   -TALFRRSLFVTRS 300
 a4 AFDTDYRIVDFYY-SKFAS-EN  34   FTLYYFIIAG 220   LQKLLRRTTNYWELWS 253   -TALYRRSLFVTKS 277
 a5 LYDSNFTIREIYW-P-LTT-TN  33   --LYIWLVAG 205   PAILFKRVKDYWRAWL 238   ---ILRRSLLILQS 260
 a6 LADSKLNYRELYY-P-LPI-DN  33   TVFYCWLVCG 203   SRELLRRTENYWKAWL 236   YDSLVRRSLLIIAA 255
 a7 KIGDYGEIEYLFY-PQVGY-ET  35   FNIYILPQRF 215   SENIKNLSMNYWKHII 248   IYSITKRALMTLLM 278
 b1 VLDRYGELHRLWW-PQK---DY  37   KCMTLFFAFG 216   FERLRSQTLADWEAYF 249   -DRLYRRSLAVFRL 275
 b2 FTLSHGILNEVYY-PRI---DS  55   DTATLALGFG 239   FRRAHDAYIANWRGWQ 272   VNS-YRVSTAVLAA 298
 b3 FTISHGIL-EVYY-PRL---DS  54   TKAVLALGFG 238   FEPAAKTYLQNWREWQ 271   INT-YRTSTAVLAT 297
 b4 YTLTEGTMSEVYY-PHA---DT  80   TEFSLALGFG 263   YKKVSKSYTGEWKKYL 296   LRTQYDVSLMTVKS 323
 b5 FTLANGAISEVYY-PTI---DT  69   SEFEIVLSFG 251   YNNLKNNYIDEWTKYC 284   ANSLYYNSMMILKA 307
    ssss--ssssss---------s       ssssssssss      hhhhhhhhhhhhhhhh        -hhhhhhhhhhhh
    βS3----βS4---------βS5       --βS18----      ------LH2-------        ------αH1-----
```

(b)

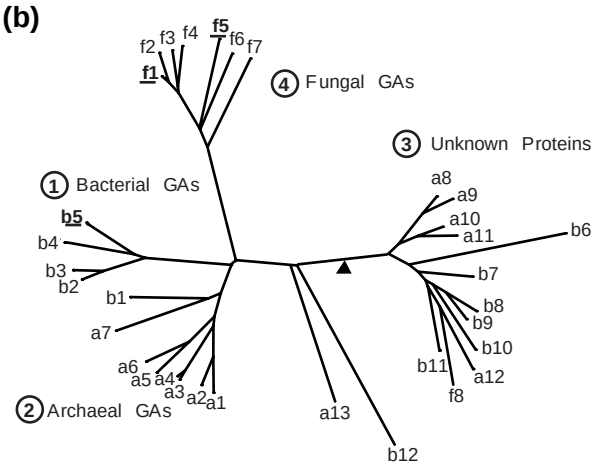


Figure 4

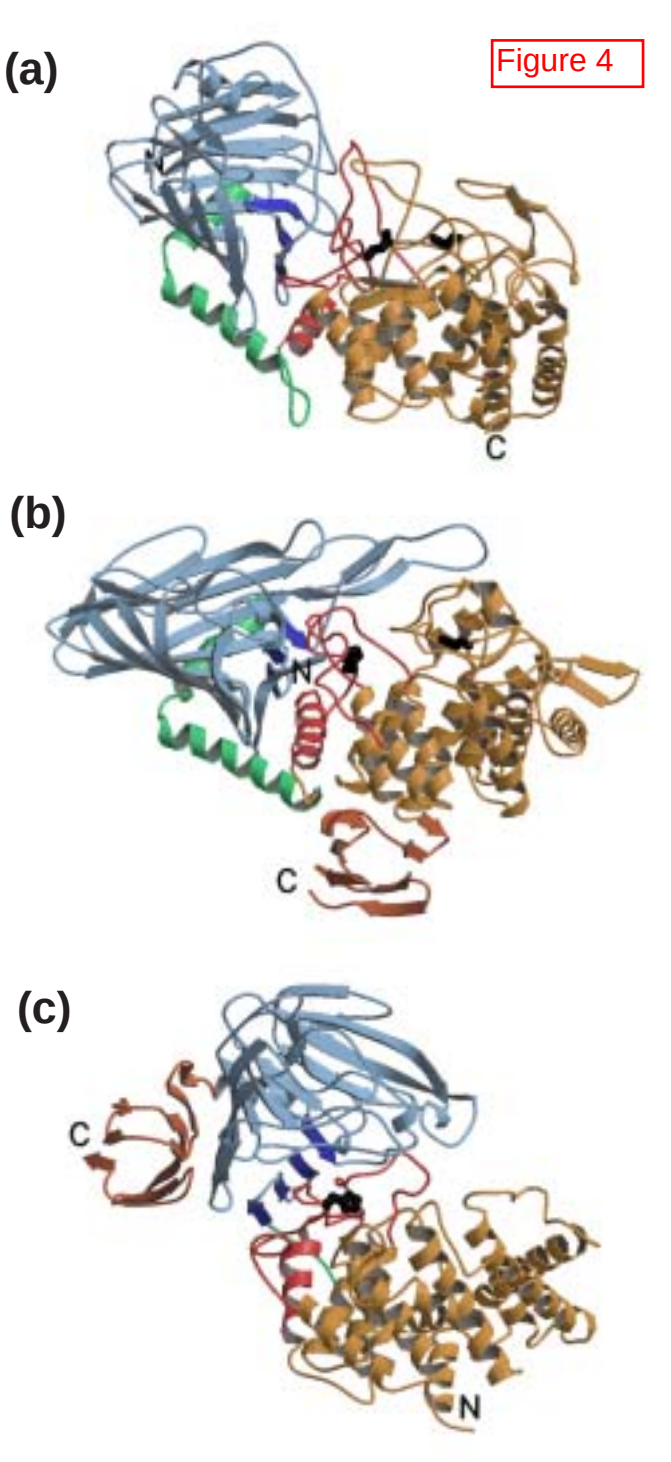


Figure 5

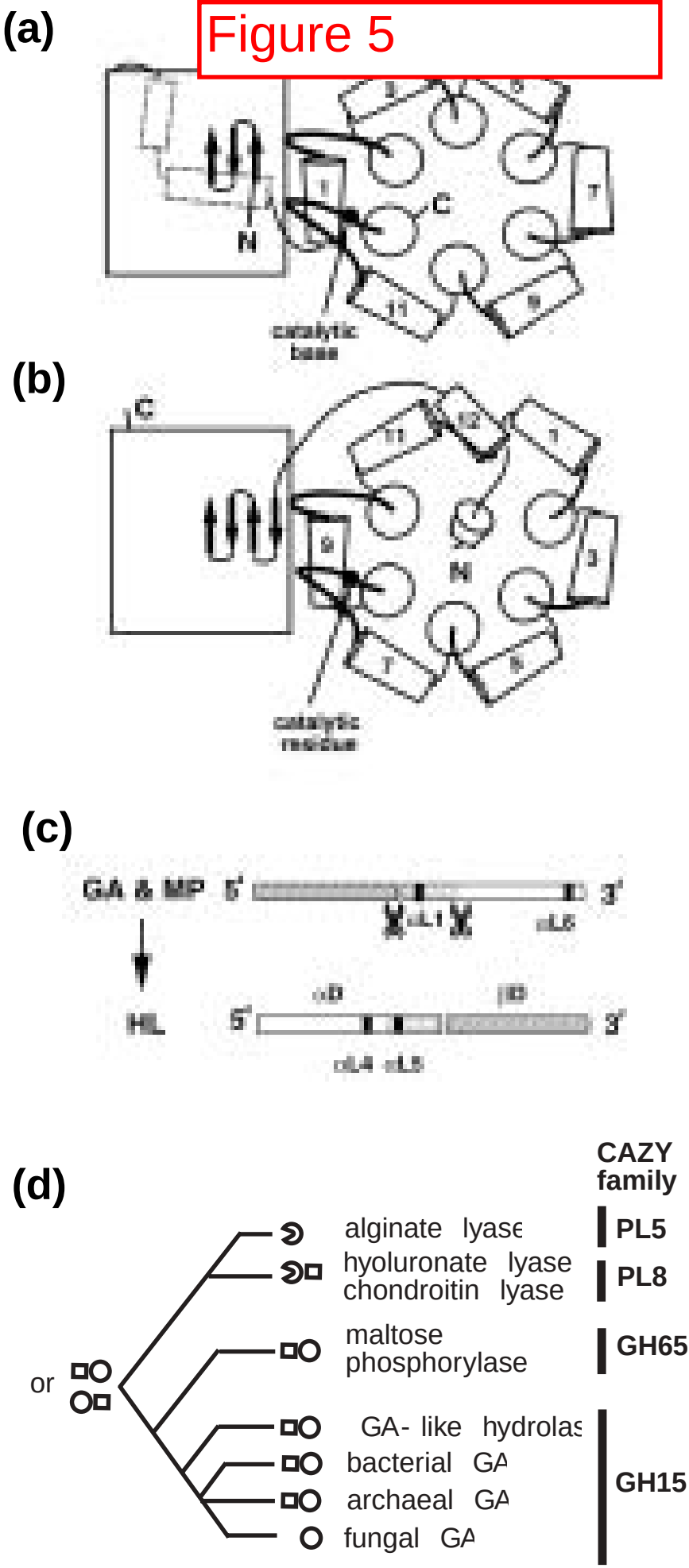

(a)
N
C
catalytic base
(b)
C
N
catalytic residue
(c)
GA & MP
5'
3'
HL
(d)
CAZY family
alginate lyase
PL5
hyoluronate lyase
chondroitin lyase
PL8
maltose phosphorylase
GH65
or
GA- like hydrolas
bacterial GA
archaeal GA
fungal GA
GH15

Figure 6

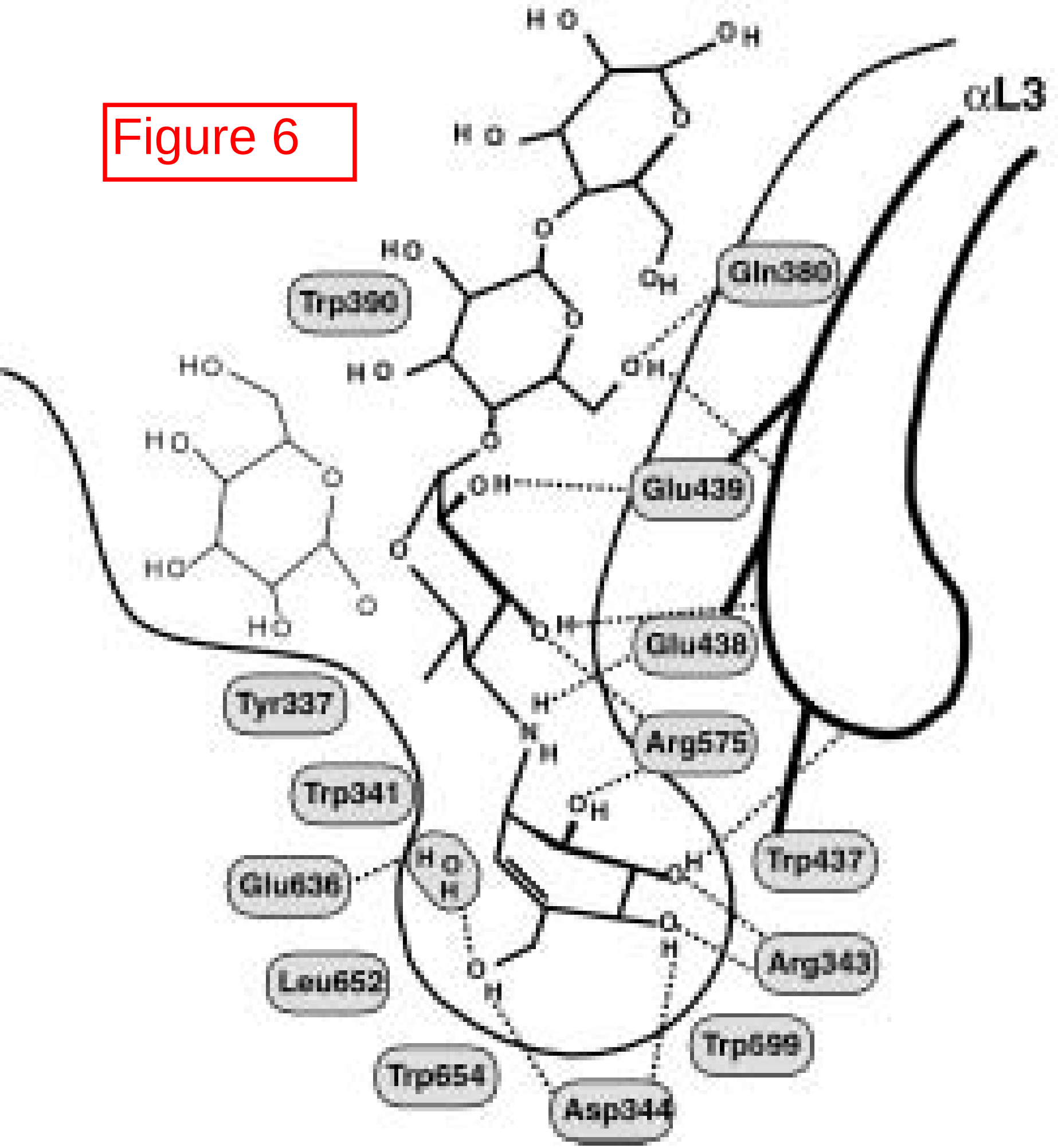
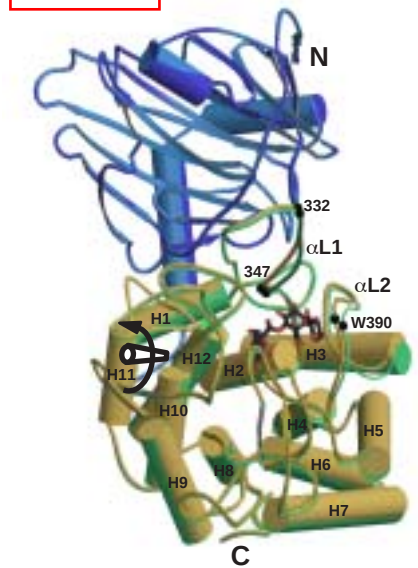
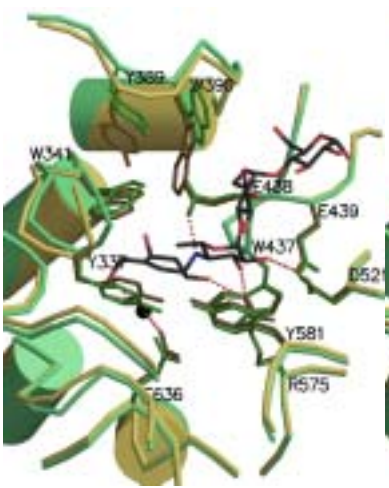

αL3
Gln389
Trp390
Glu439
Glu438
Tyr337
Arg575
Trp341
Trp437
Glu636
Arg343
Leu652
Trp659
Trp654
Asp344

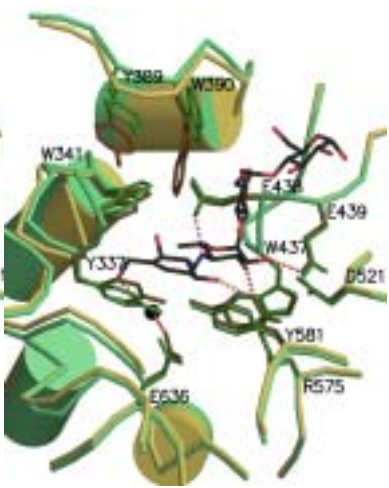
Figure 7

(a)

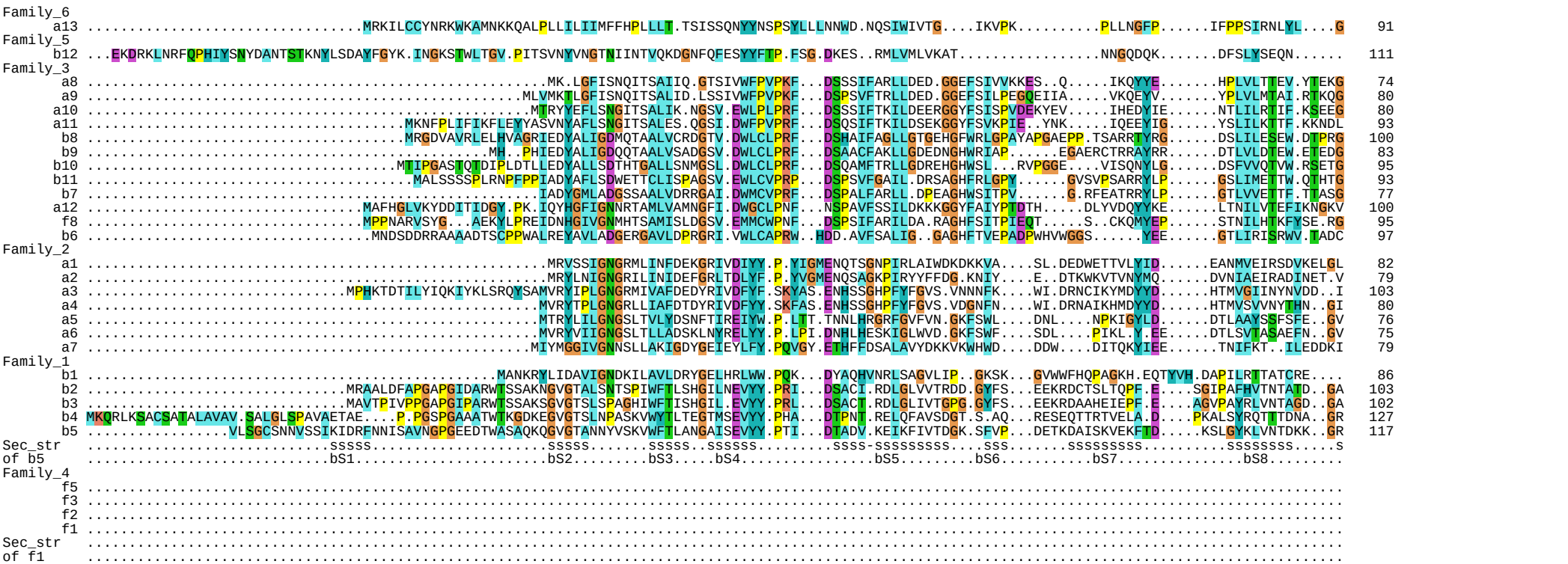

N
332
αL1
347
αL2
H1
W390
H12
H3
H11
H2
H10
H4
H5
H8
H6
H9
H7
C

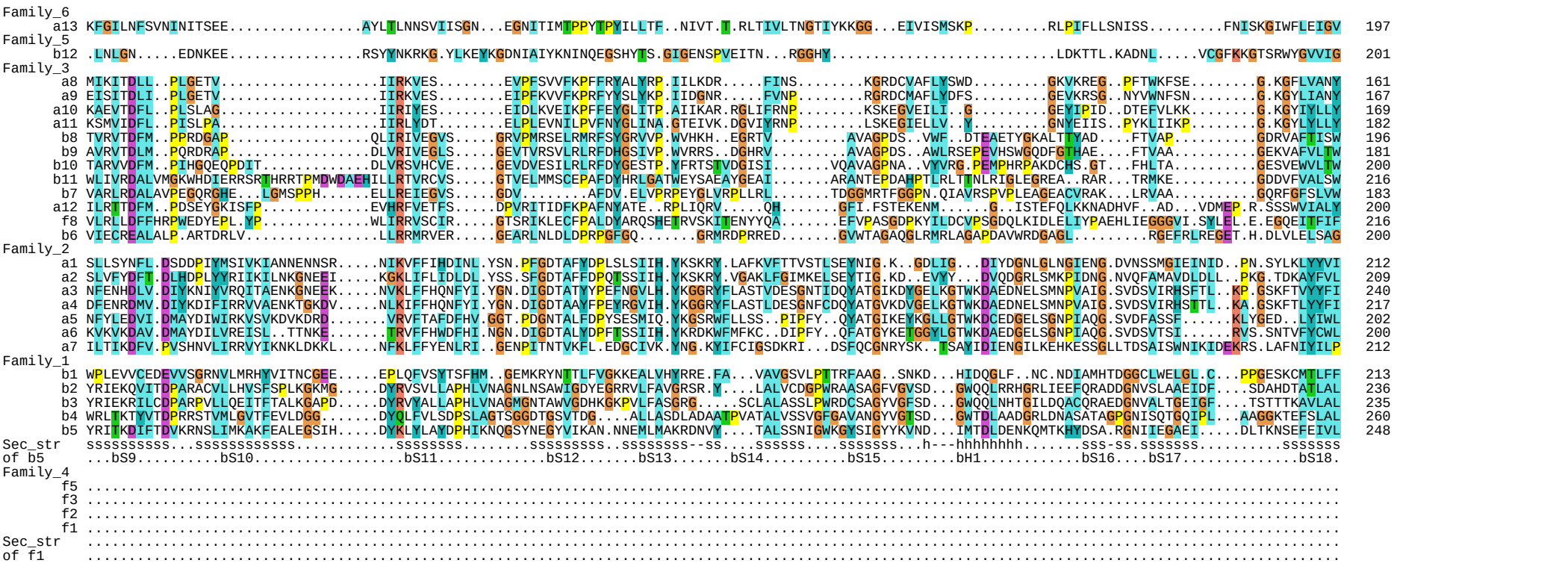
(b)

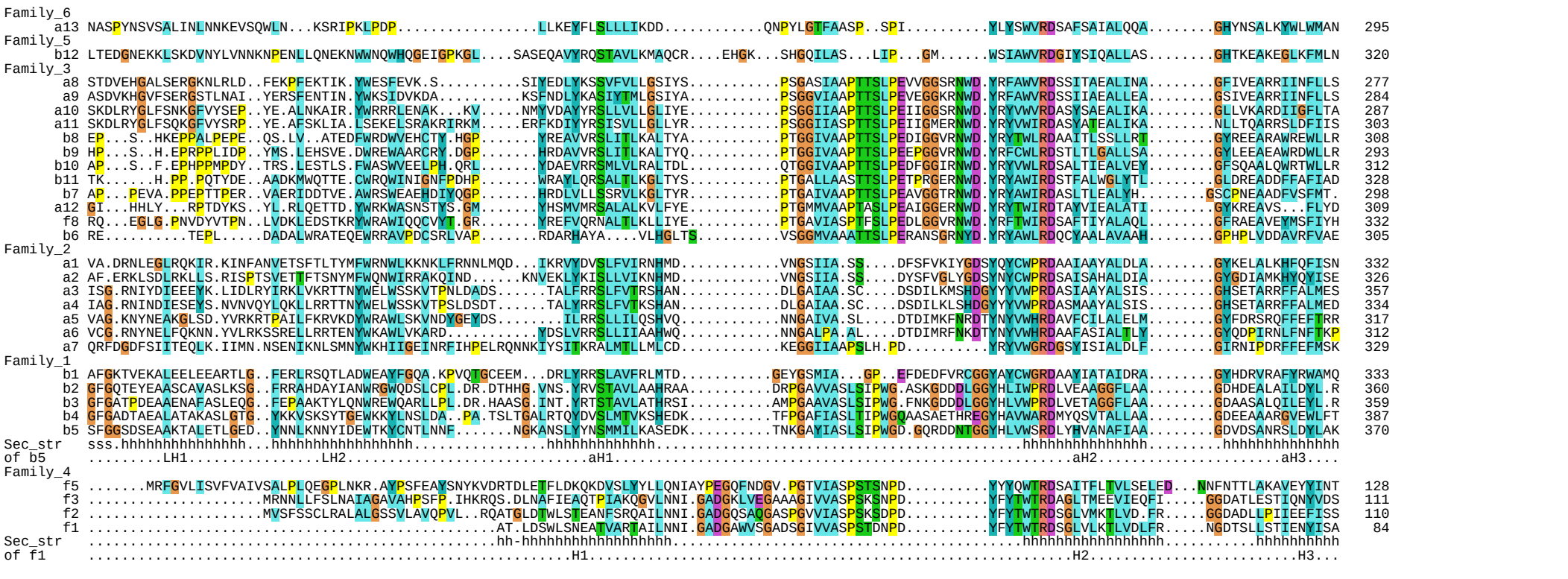

Y389
W390
W341
E438
E439
W437
Y337
D521
Y581
R575
E636

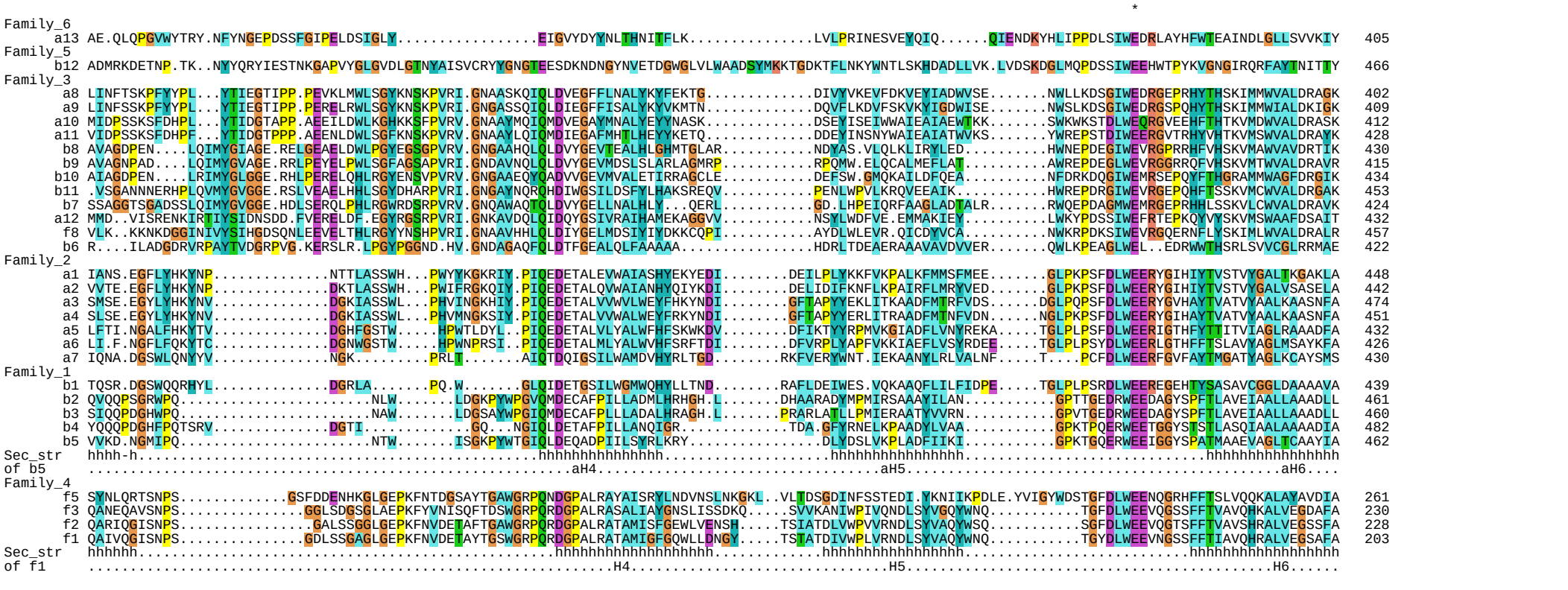

Y389
W390
W341
E438
E439
W437
Y337
D521
Y581
R575
E636

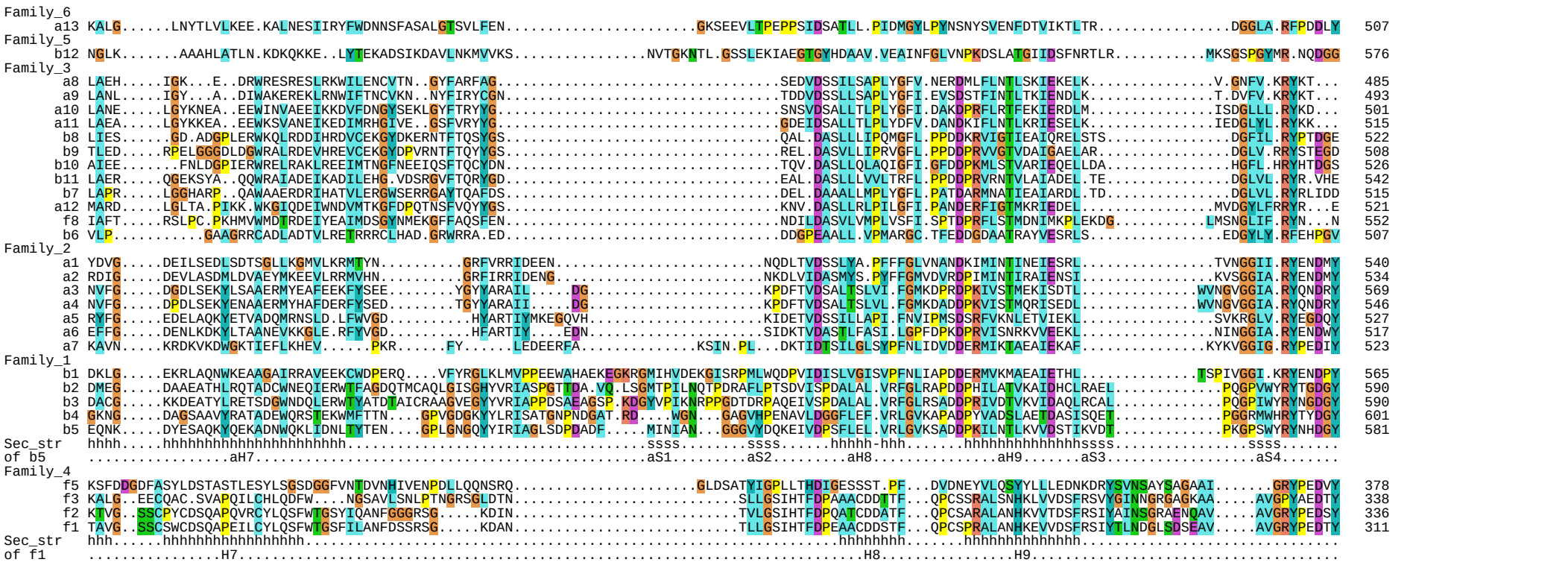
Supplemental Materials

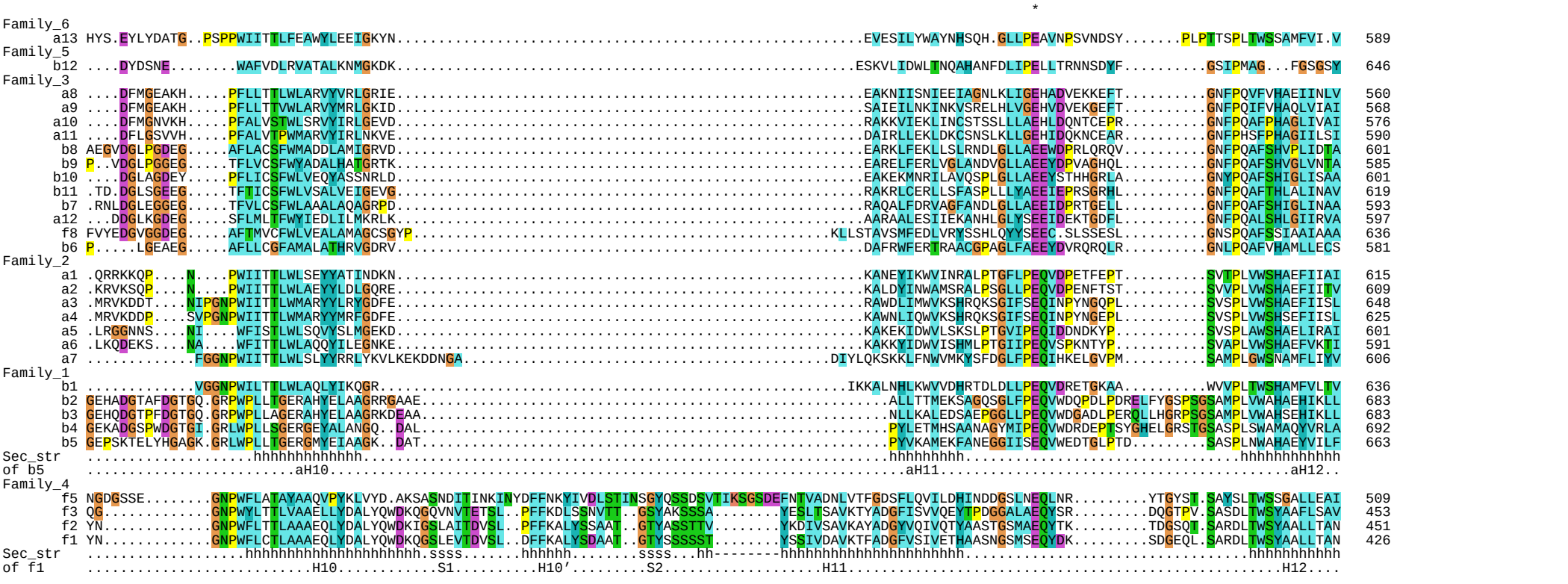



Family_6
a13 KFGILNFSVNINITSEE.....AYLTLNNSVIISGN...EGNITIMTPPYTPYILLTF..NIVT.T.RLTIVLTNGTIYKKGG...EIVISMSKP.....RLPIFLLSNISS.....FNISKGIWFLEIGV 197
Family_5
b12 .LNLGN.....EDNKEE.....RSYYNKRKG.YLKEYKGDNIAIYKNINQEGSHYTS.GIGENSPVEITN...RGGHY.....LDKTTL.KADNL.....VCGFKKGTSRWYGVVIG 201
Family_3
a8 MIKITDLL..PLGETV.....IIRKVES.....EVPFSVVFKPFFRYALYRP.IILKDR.....FINS.....KGRDCVAFLYSWD.....GKVKREG..PFTWKFSE.....G.KGFLVANY 161
a9 EISITDLI..PLGETV.....IIRKVES.....EIPFKVVFKPRFYYSLYKP.IIDGNR.....FVNP.....RGRDCMAFLYDFS.....GEVKRSG..NYVWNFSN.....G.KGYLIANY 167
a10 KAEVTDFL..PLSLAG.....IIRIYES.....EIDLKVEIKPFFEYGLITP.AIIKAR.RGLIFRNP.....KSKEGVEILI..G.....GEYIPID..DTEFVLKK.....G.KGYIYLLY 169
a11 KSMVIDFL..PISLPA.....IIRLYDT.....ELPLEVNILPVFNYGLINA.GTEIVK.DGVIYRNP.....LSKEGIELLV..Y.....GNYEIIS..PYKLIIKP.....G.KGYLYLLY 182
b8 TVRVTDFM..PPRDGAP.....QLIRIVEGVS.....GRVPMRSELRMRFSYGRVVP.WVHKH..EGRTV.....AVAGPDS..VWF..DTEAETYGKALTTYAD....FTVAP.....GDRVAFTISW 196
b9 AVRVTDLM..PQRDRAP.....DLVRIVEGLS.....GEVTVRSVLRLRFDHGSIVP.WVRRS..DGHRV.....AVAGPDS..AWLRSEPEVHSWGQDFGTHAE....FTVAA.....GEKVAFVLTW 181
b10 TARVVDFM..PIHGQEQPDIT.....DLVRSVHCVE.....GEVDVESILRLRFDYGESTP.YFRTSTVDGISI.....VQAVAGPNA..VYVRG.PEMPHRPAKDCHS.GT...FHLTA.....GESVEWVLTW 200
b11 WLIVRDALVMGKWHDIERRSRTHRRTPMDWDAEHILLRTVRCVS.....GTVELMMSCEPAFDYHRLGATWEYSAEAYGEAI.....ARANTEPDAHPTLRLTTNLRIGLEGREA..RAR...TRMKE.....GDDVFVALSW 216
b7 VARLRDALAVPEGQRGHE...LGMSPPH.....ELLREIEGVS.....GDV.....AFDV.ELVPRPEYGLVRPLLRL.....TDGGMRTFGGPN.QIAVRSPVPLEAGEACVRAK...LRVAA.....GQRFGFSLVW 183
a12 ILRTTDFM..PDSEYGKISFP.....EVHRFVETFS.....DPVRITIDFKPAFNYATE..RPLIQRV.....QH.....GFI.FSTEKENM.....G..ISTEFQLKKNADHVF..AD...VDMEP.R.SSSWVIALY 200
f8 VLRLLDFFHRPWEDYEPL.YP.....WLIRRVSCIR.....GTSRIKLECFPALDYARQSHETRVSKITENYYQA.....EFVPASGDPKYILDCVPSGDQLKIDLELIYPAEHLIEGGGVI.SYLEL.E.EGQEITFIF 216
b6 VIECREALALP.ARTDRLV.....LLRRMRVER.....GEARLNLDLDPRPGFGQ.....GRMRDPRRED.....GVWTAGAQGLRMRLAGAPDAVWRDGAGL.....RGEFRLREGET.H.DLVLELSAG 200
Family_2
a1 SLLSYNFL.DSDDPIYMSIVKIANNENNSR.....NIKVFFIHDINL.YSN.PFGDTAFYDPLSLSIIH.YKSKRY.LAFKVFTTVSTLSEYNIG.K..GDLIG...DIYDGNLGLNGIENG.DVNSSMGIEINID..PN.SYLKLYYVI 212
a2 SLVFYDFT.DLHDPLYYRIIKILNKGNEEI.....KGKLIFLIDLDL.YSS.SFGDTAFFDPQTSSIIH.YKSKRY.VGAKLFGIMKELSEYTIG.KD..EVYY...DVQDGRLSMKPIDNG.NVQFAMAVDLDLL..PKG.TDKAYFVL 209
a3 NFENHDLV.DIYKNIYVRQITAENKGNEEK.....NVKLFFHQNFYI.YGN.DIGDTATYYPEFNGVLH.YKGGRYFLASTVDESGNTIDQYATGIKDYGELKGTWKDAEDNELSMNPVAIG.SVDSVIRHSFTL..KP.GSKFTVYYFI 240
a4 DFENRDMV.DIYKDIFIRRVVAENKTGKDV.....NLKIFFHQNFYI.YGN.DIGDTAAYFPEYRGVIH.YKGGRYFLASTLDESGNFCDQYATGVKDVGELKGTWKDAEDNELSMNPVAIG.SVDSVIRHSTTL..KA.GSKFTLYYFI 217
a5 NFYLEDVI.DMAYDIWIRKVSVKDVKDRD.....VRVFTAFDFHV.GGT.PDGNTALFDPYSESMIQ.YKGSRWFLLSS..PIPFY..QYATGIKEYKGLLGTWKDCEDGELSGNPIAQG.SVDFASSF.....KLYGED..LYIWL 202
a6 KVKVKDAV.DMAYDILVREISL..TTNKE.....TRVFFHWDFHI.NGN.DIGDTALYDPFTSSIIH.YKRDKWFMFKC..DIPFY..QFATGYKETGGYLGTWKDAEDGELSGNPIAQG.SVDSVTSI.....RVS.SNTVFYCWL 200
a7 ILTIKDFV.PVSHNVLIRRVYIKNKLDKKL.....NFKLFFYENLRI..GENPITNTVKFL.EDGCIVK.YNG.KYIFCIGSDKRI...DSFQCGNRYSK..TSAYIDIENGILKEHKESSGLLTDSAISWNIKIDEKRS.LAFNIYILP 212
Family_1
b1 WPLEVVCEDEVVSGRNVLMRHYVITNCGEE.....EPLQFVSYTSFHM..GEMKRYNTTLFVGKKEALVHYRRE.FA...VAVGSVLPTTRFAAG..SNKD...HIDQGLF..NC.NDIAMHTDGGCLWELGL.C...PPGESKCMTLFF 213
b2 YRIEKQVITDPARACVLLHVSFSPLKGKMG.....DYRVSVLLAPHLVNAGNLNSAWIGDYEGRRVLFAVGRSR.Y....LALVCDGPWRAASAGFVGVSD...GWQQLRRHGRLIEEFQRADDGNVSLAAEIDF....SDAHDTATLAL 236
b3 YRIEKRILCDPARPVLLQEITFTALKGAPD.....DYRVYALLAPHLVNAGMGNTAWVGDHKGKPVLFASGRG.....SCLALASSLPWRDCSAGYVGFSD...GWQQLNHTGILDQACQRAEDGNVALTGEIGF....TSTTTKAVLAL 235
b4 WRLTKTYVTDPRRSTVMLGVTFEVLDGG.....DYQLFVLSDPSLAGTSGGDTGSVTDG....ALLASDLADAATPVATALVSSVGFGAVANGYVGTSD...GWTDLAADGRLDNASATAGPGNISQTGQIPL...AAGGKTEFSLAL 260
b5 YRITKDIFTDVKRNSLIMKAKFEALEGSIH.....DYKLYLAYDPHIKNQGSYNEGYVIKAN.NNEMLMAKRDNVY....TALSSNIGWKGYSIGYYKVND...IMTDLDENKQMTKHYDSA.RGNIIEGAEI.....DLTKNSEFEIVL 248
Sec_str sssssssss...sssssssssss.....sssssss.....sssssss..sssssss--ss....ssssss....sssssss...h---hhhhhhhh.....sss-ss.sssssss.....sssssss
of b5 ...bS9.....bS10.....bS11.....bS12.....bS13.....bS14.....bS15.....bH1.....bS16....bS17.....bS18.
Family_4
f5 ....
f3 ....
f2 ....
f1 ....
Sec_str ....
of f1 ....


Family_6
a13 NASPYNSVSALINLNNKEVSQWLN...KSRIPKLPDP.....LLKEYFLSLLLIKDD.....QNPYLGTFAASP..SPI.....YLYSWVRDSAFSAIALQQA.....GHYNSALKYWLWMAN 295
Family_5
b12 LTEDGNEKKLSKDVNYLVNNKNPENLLQNEKNWWNQWHQGEIGPKGL....SASEQAVYRQSTAVLKMAQCR....EHGK...SHGQILAS...LIP...GM.....WSIAWVRDGIYSIQALLAS.....GHTKEAKEGLKFMLN 320
Family_3
a8 STDVEHGALSERGKNLRLD..FEKPFEKTIK.YWESFEVK.S.....SIYEDLYKSSVFVLLGSIYS.....PSGASIAAPTTSLPEVVGGSRNWD.YRFAWVRDSSITAEALINA.....GFIVEARRIINFLLS 277
a9 ASDVKHGVFSERGSTLNAI..YERSFENTIN.YWKSIDVKDA.....KSFNDLYKASIYTMLGSIYA.....PSGGVIAAPTTSLPEVEGGKRNWD.YRFAWVRDSSIIAEALLEA.....GSIVEARRIINFLLS 284
a10 SKDLRYGLFSNKGFVYSEP..YE.ALNKAIR.YWRRRLENAK...KV.....NMYVDAYYRSLLVLLGLIYE.....PSGGIIAAPTTSLPEIIGGSRNWD.YRYVWVRDASYSAEALIKA.....GLLVKARDIIGFLTA 287
a11 SKDLRYGLFSQKGFVYSRP..YE.AFSKLIA.LSEKELSRAKRIRKM.....ERFKDIYYRSISVLLGLLYR.....PSGGIIASPTTSLPEIIGMERNWD.YRYVWIRDASYATEALIKA.....NLLTQARRSLDFIIS 303
b8 EP...S..HKEPPALPEPE..QS.LV..ATEDFWRDWVEHCTY.HGP.....YREAVVRSLITLKALTYA.....PTGGIVAAPTTSLPEDIGGVRNWD.YRYTWLRDAAITLSSLLRT.....GYREEARAWREWLLR 308
b9 HP...S..H.EPRPPLIDP..YMS.LEHSVE.DWREWAARCRY.DGP.....HRDAVVRSLITLKALTYQ.....PTGGIVAAPTTSLPEEPGGVRNWD.YRFCWLRDSTLTLGALLSA.....GYLEEAEAWRDWLLR 293
b10 AP...S..F.EPHPPMPDY..TRS.LESTLS.FWASWVEELPH.QRL.....YDAEVRRSMLVLRALTDL.....QTGGIVAAPTTSLPEDFGGIRNWD.YRYVWLRDSALTIEALVEY.....GFSQAALQWRTWLLR 312
b11 TK.....H.PP.PQTYDE..AADKMWQTTE.CWRQWINIGNFPDHP.....WRAYLQRSALTLKGLTYS.....PTGALLAASTTSLPETPRGERNWD.YRYAWIRDSTFALWGLYTL.....GLDREADDFFAFIAD 328
b7 AP...PEVA.PPEPTTPER..VAERIDDTVE.AWRSWEAEHDIYQGP.....HRDLVLLSSRVLKGLTYR.....PTGAIVAAPTTSLPEAVGGTRNWD.YRYAWIRDASLTLEALYH.....GSCPNEAADFVSFMT. 298
a12 GI...HHLY...RPTDYKS..YL.RLQETTD.YWRKWASNSTYS.GM.....YHSMVMRSALALKVLFYE.....PTGMMVAAPTASLPEAIGGERNWD.YRYTWIRDTAYVIEALATI.....GYKREAVS...FLYD 309
f8 RQ...EGLG.PNVDYVTPN..LVDKLEDSTKRYWRAWIQQCVYT.GR.....YREFVQRNALTLKLLIYE.....PTGAVIASPTFSLPEDLGGVRNWD.YRFTWIRDSAFTIYALAQL.....GFRAEAVEYMSFIYH 332
b6 RE.....TEPL.....DADALWRATEQEWRRAVPDCSRLVAP.....RDARHAYA....VLHGLTS.....VSGGMVAAATTSLPERANSGRNYD.YRYAWLRDQCYAALAVAAH.....GPHPLVDDAVRFVAE 305
Family_2
a1 VA.DRNLEGLRQKIR.KINFANVETSFTLTYMFWRNWLKKNKLFRNNLMQD...IKRVYDVSLFVIRNHMD.....VNGSIIA.SS....DFSFVKIYGDSYQYCWPRDAAIAAYALDLA.....GYKELALKHFQFISN 332
a2 AF.ERKLSDLRKLLS.RISPTSVETTFTSNYMFWQNWIRRAKQIND.....KNVEKLYKISLLVIKNHMD.....VNGSIIA.SS....DYSFVGLYGDSYNYCWPRDSAISAHALDIA.....GYGDIAMKHYQYISE 326
a3 ISG.RNIYDIEEEYK.LIDLRYIRKLVKRTTNYWELWSSKVTPNLDADS.....TALFRRSLFVTRSHAN.....DLGAIAA.SC....DSDILKMSHDGYYYVWPRDASIAAYALSIS.....GHSETARRFFALMES 357
a4 IAG.RNINDIESEYS.NVNVQYLQKLLRRTTNYWELWSSKVTPSLDSDT.....TALYRRSLFVTKSHAN.....DLGAIAA.SC....DSDILKLSHDGYYYVWPRDASMAAYALSIS.....GHSETARRFFALMED 334
a5 VAG.KNYNEAKGLSD.YVRKRTPAILFKRVKDYWRAWLSKVNDYGEYDS.....ILRRSLLILQSHVQ.....NNGAIVA.SL....DTDIMKFNRDTYNYVWHRDAVFCILALELM.....GYFDRSRQFFEFTRR 317
a6 VCG.RNYNELFQKNN.YVLRKSSRELLRRTENYWKAWLVKARD.....YDSLVRRSLLIIAAHWQ.....NNGALPA.AL....DTDIMRFNKDTYNYVWHRDAAFASIALTLY.....GYQDPIRNLFNFTKP 312
a7 QRFDGDFSITEQLK.IIMN.NSENIKNLSMNYWKHIIGEINRFIHPELRQNNKIYSITKRALMTLLMLCD.....KEGGIIAAPSLH.PD.....YRYVWGRDGSYISIALDLF.....GIRNIPDRFFEFMSK 329
Family_1
b1 AFGKTVEKALEELEEARTLG..FERLRSQTLADWEAYFGQA.KPVQTGCEEM...DRLYRRSLAVFRLMTD.....GEYGSMIA...GP..EFDEDFVRCGGYAYCWGRDAAYIATAIDRA.....GYHDRVRAFYRWAMQ 333
b2 GFGQTEYEAASCAVASLKSG..FRRAHDAYIANWRGWQDSLCPL.DR.DTHHG.VNS.YRVSTAVLAAHRAA.....DRPGAVVASLSIPWG.ASKGDDDLGGYHLIWPRDLVEAAGGFLAA.....GDHDEALAILDYL.R 360
b3 GFGATPDEAAENAFASLEQG..FEPAAKTYLQNWREWQARLLPL.DR.HAASG.INT.YRTSTAVLATHRSI.....AMPGAAVASLSIPWG.FNKGDDDLGGYHLVWPRDLVETAGGFLAA.....GDAASALQILEYL.R 359
b4 GFGADTAEALATAKASLGTG..YKKVSKSYTGEWKKYLNSLDA..PA.TSLTGALRTQYDVSLMTVKSHEDK.....TFPGAFIASLTIPWGQAASAETHREGYHAVWARDMYQSVTALLAA.....GDEEAAARGVEWLFT 387
b5 SFGGSDSEAAKTALETLGED..YNNLKNNYIDEWTKYCNTLNNF.....NGKANSLYYNSMMILKASEDK.....TNKGAYIASLSIPWGD.GQRDDNTGGYHLVWSRDLYHVANAFIAA.....GDVDSANRSLDYLAK 370
Sec_str sss.hhhhhhhhhhhhhh...hhhhhhhhhhhhhhhhh.....hhhhhhhhhhhhh.....hhhhhhhhhhhhhhh.....hhhhhhhhhhhhh
of b5 .....LH1.....LH2.....aH1.....aH2.....aH3....
Family_4
f5 .....MRFGVLISVFVAIVSALPLQEGPLNKR.AYPSFEAYSNYKVDRTDLETFLDKQKDVSLYYLLQNIAYPEGQFNDGV.PGTVIASPSTSNPD.....YYYQWTRDSAITFLTVLSELED...NNFNTTLAKAVEYYINT 128
f3 .....MRNNLLFSLNAIAGAVAHPSFP.IHKRQS.DLNAFIEAQTPIAKQGVLNNI.GADGKLVEGAAAGIVVASPSKSNPD.....YFYTWTRDAGLTMEEVIEQFI.....GGDATLESTIQNYVDS 111
f2 .....MVSFSSCLRALALGSSVLAVQPVL..RQATGLDTWLSTEANFSRQAILNNI.GADGQSAQGASPGVVIASPSKSDPD.....YFYTWTRDSGLVMKTLVD.FR.....GGDADLLPIIEEFISS 110
f1 .....AT.LDSWLSNEATVARTAILNNI.GADGAWVSGADSGIVVASPSTDNPD.....YFYTWTRDSGLVLKTLVDLFR.....NGDTSLLSTIENYISA 84
Sec_str .....hh-hhhhhhhhhhhhhhhh.....hhhhhhhhhhhhhhhhh.....hhhhhhhhhh
of f1 .....H1.....H2.....H3...


*
Family_6
a13 AE.QLQPGVWYTRY.NFYNGEPDSSFGIPELDSIGLY.....EIGVYDYYNLTHNITFLK.....LVLPRINESVEYQIQ.....QIENDKYHLIPPDLSIWEDRLAYHFWTEAINDLGLLSVVKIY 405
Family_5
b12 ADMRKDETNP.TK..NYYQRYIESTNKGAPVYGLGVDLGTNYAISVCRYYGNGTEESDKNDNGYNVETDGWGLVLWAADSYMKKTGDKTFLNKYWNTLSKHDADLLVK.LVDSKDGLMQPDSSIWEEHWTPYKVGNGIRQRFAYTNITTY 466
Family_3
a8 LINFTSKPFYYPL...YTIEGTIPP.PEVKLMWLSGYKNSKPVRI.GNAASKQIQLDVEGFFLNALYKYFEKTG.....DIVYVKEVFDKVEYIADWVSE.....NWLLKDSGIWEDRGEPRHYTHSKIMMWVALDRAGK 402
a9 LINFSSKPFYYPL...YTIEGTIPP.PERELRWLSGYKNSKPVRI.GNGASSQIQLDIEGFFISALYKYVKMTN.....DQVFLKDVFSKVKYIGDWISE.....NWSLKDSGIWEDRGSPQHYTHSKIMMWIALDKIGK 409
a10 MIDPSSKSFDHPL...YTIDGTAPP.AEEILDWLKGHKKSFPVRV.GNAAYMQIQMDVEGAYMNALYEYYNASK.....DSEYISEIWWAIEAIAEWTKK.....SWKWKSTDLWEQRGVEEHFTHTKVMDWVALDRASK 412
a11 VIDPSSKSFDHPF...YTIDGTPPP.AEENLDWLSGFKNSKPVRV.GNAAYLQIQMDIEGAFMHTLHEYYKETQ.....DDEYINSNYWAIEAIATWVKS.....YWREPSTDIWEERGVTRHYVHTKVMSWVALDRAYK 428
b8 AVAGDPEN....LQIMYGIAGE.RELGEAELDWLPGYEGSGPVRV.GNGAAHQLQLDVYGEVTEALHLGHMTGLAR.....NDYAS.VLQLKLIRYLED.....HWNEPDEGIWEVRGPRRHFVHSKVMAWVAVDRTIK 430
b9 AVAGNPAD....LQIMYGVAGE.RRLPEYELPWLSGFAGSAPVRI.GNDAVNQLQLDVYGEVMDSLSLARLAGMRP.....RPQMW.ELQCALMEFLAT.....AWREPDEGLWEVRGGRRQFVHSKVMTWVALDRAVR 415
b10 AIAGDPEN....LRIMYGLGGE.RHLPERELQHLRGYENSVPVRV.GNGAAEQYQADVVGEVMVALETIRRAGCLE.....DEFSW.GMQKAILDFQEA.....NFDRKDQGIWEMRSEPQYFTHGRAMMWAGFDRGIK 434
b11 .VSGANNNERHPLQVMYGVGGE.RSLVEAELHHLSGYDHARPVRI.GNGAYNQRQHDIWGSILDSFYLHAKSREQV.....PENLWPVLKRQVEEAIK.....HWREPDRGIWEVRGEPQHFTSSKVMCWVALDRGAK 453
b7 SSAGGTSGADSSLQIMYGVGGE.HDLSERQLPHLRGWRDSRPVRV.GNQAWAQTQLDVYGELLNALHLY...QERL.....GD.LHPEIQRFAAGLADTALR.....RWQEPDAGMWEMRGEPRHHLSSKVLCWVALDRAVK 424
a12 MMD..VISRENKIRTIYSIDNSDD.FVERELDF.EGYRGSRPVRI.GNKAVDQLQIDQYGSIVRAIHAMEKAGGVV.....NSYLWDFVE.EMMAKIEY.....LWKYPDSSIWEFRTEPKQYVYSKVMSWAAFDSAIT 432
f8 VLK..KKNKDGGINIVYSIHGDSQNLEEVELTHLRGYYNSHPVRI.GNAAVHHLQLDIYGELMDSIYIYDKKCQPI.....AYDLWLEVR.QICDYVCA.....NWKRPDKSIWEVRGQERNFLYSKIMLWVALDRALR 457
b6 R....ILADGDRVRPAYTVDGRPVG.KERSLR.LPGYPGGND.HV.GNDAGAQFQLDTFGEALQLFAAAAA.....HDRLTDEAERAAAVAVDVVER.....QWLKPEAGLWEL..EDRWWTHSRLSVVCGLRRMAE 422
Family_2
a1 IANS.EGFLYHKYNP.....NTTLASSWH...PWYYKGKRIY.PIQEDETALEVWAIASHYEKYEDI.....DEILPLYKKFVKPALKFMMSFMEE.....GLPKPSFDLWEERYGIHIYTVSTVYGALTKGAKLA 448
a2 VVTE.EGFLYHKYNP.....DKTLASSWH...PWIFRGKQIY.PIQEDETALQVWAIANHYQIYKDI.....DELIDIFKNFLKPAIRFLMRYVED.....GLPKPSFDLWEERYGIHIYTVSTVYGALVSASELA 442
a3 SMSE.EGYLYHKYNV.....DGKIASSWL...PHVINGKHIY.PIQEDETALVVWVLWEYFHKYNDI.....GFTAPYYEKLITKAADFMTRFVDS.....DGLPQPSFDLWEERYGVHAYTVATVYAALKAASNFA 474
a4 SLSE.EGYLYHKYNV.....DGKIASSWL...PHVMNGKSIY.PIQEDETALVVWALWEYFRKYNDI.....GFTAPYYERLITRAADFMTNFVDN.....NGLPKPSFDLWEERYGIHAYTVATVYAALKAASNFA 451
a5 LFTI.NGALFHKYTV.....DGHFGSTW.....HPWTLDYL..PIQEDETALVLYALWFHFSKWKDV.....DFIKTYYRPMVKGIADFLVNYREKA.....TGLPLPSFDLWEERIGTHFYTTITVIAGLRAAADFA 432
a6 LI.F.NGFLFQKYTC.....DGNWGSTW.....HPWNPRSI..PIQEDETALMLYALWVHFSRFTDI.....DFVRPLYAPFVKKIAEFLVSYRDEE.....TGLPLPSYDLWEERLGTHFFTSLAVYAGLMSAYKFA 426
a7 IQNA.DGSWLQNYYV.....NGK.....PRLT.....AIQTDQIGSILWAMDVHYRLTGD.....RKFVERYWNT.IEKAANYLRLVALNF.....T....PCFDLWEERFGVFAYTMGATYAGLKCAYSMS 430
Family_1
b1 TQSR.DGSWQQRHYL.....DGRLA.....PQ.W.....GLQIDETGSILWGMWQHYLLTND.....RAFLDEIWES.VQKAAQFLILFIDPE.....TGLPLPSRDLWEEREGEHTYSASAVCGGLDAAAAVA 439
b2 QVQQPSGRWPQ.....NLW.....LDGKPYWPGVQMDECAFPILLADMLHRHGH.L.....DHAARADYMPMIRSAAAYILAN.....GPTTGEDRWEEDAGYSPFTLAVEIAALLAAADLL 461
b3 SIQQPDGHWPQ.....NAW.....LDGSAYWPGIQMDECAFPLLLADALHRAGH.L.....PRARLATLLPMIERAATYVVRN.....GPVTGEDRWEEDAGYSPFTLAVEIAALLAAADLL 460
b4 YQQQPDGHFPQTSRV.....DGTI.....GQ...NGIQLDETAFPILLANQIGR.....TDA.GFYRNELKPAADYLVAA.....GPKTPQERWEETGGYSTSTLASQIAALAAAADIA 482
b5 VVKD.NGMIPQ.....NTW.....ISGKPYWTGIQLDEQADPIILSYRLKRY.....DLYDSLVKPLADFIIKI.....GPKTGQERWEEIGGYSPATMAAEVAGLTCAAYIA 462
Sec_str hhhh-h.....hhhhhhhhhhhhhh.....hhhhhhhhhhhhhhh.....hhhhhhhhhhhhhhh
of b5 .....aH4.....aH5.....aH6....
Family_4
f5 SYNLQRTSNPS.....GSFDDENHKGLGEPKFNTDGSAYTGAWGRPQNDGPALRAYAISRYLNDVNSLNKGKL..VLTDSGDINFSSTEDI.YKNIIKPDLE.YVIGYWDSTGFDLWEENQGRHFFTSLVQQKALAYAVDIA 261
f3 QANEQAVSNPS.....GGLSDGSGLAEPKFYVNISQFTDSWGRPQRDGPALRASALIAYGNSLISSDKQ.....SVVKANIWPIVQNDLSYVGQYWNQ.....TGFDLWEEVQGSSFFTVAVQHKALVEGDAFA 230
f2 QARIQGISNPS.....GALSSGGLGEPKFNVDETAFTGAWGRPQRDGPALRATAMISFGEWLVENSH.....TSIATDLVWPVVRNDLSYVAQYWSQ.....SGFDLWEEVQGTSFFTVAVSHRALVEGSSFA 228
f1 QAIVQGISNPS.....GDLSSGAGLGEPKFNVDETAYTGSWGRPQRDGPALRATAMIGFGQWLLDNGY.....TSTATDIVWPLVRNDLSYVAQYWNQ.....TGYDLWEEVNGSSFFTIAVQHRALVEGSAFA 203
Sec_str hhhhhh.....hhhhhhhhhhhhhhhhh.....hhhhhhhhhhhhhhhhh.....hhhhhhhhhhhhhhhhh
of f1 .....H4.....H5.....H6.....


Family_6
a13 KALG.....LNYTLVLKEE.KALNESIIRYFWDNNSFASALGTSVLFEN.....GKSEEVLTPEPPSIDSATLL.PIDMGYLPYNSNYSVENFDTVIKTLTR.....DGGLA.RFPDDLY 507
Family_5
b12 NGLK.....AAAHLATLN.KDKQKKE..LYTEKADSIKDAVLNKMVVKS.....NVTGKNTL.GSSLEKIAEGTGYHDAAV.VEAINFGLVNPKDSLATGIIDSFNRTLR.....MKSGSPGYMR.NQDGG 576
Family_3
a8 LAEH.....IGK...E..DRWRESRESLRKWILENCVTN..GYFARFAG.....SEDVDSSILSAPLYGFV.NERDMLFLNTLSKIEKELK.....V.GNFV.KRYKT... 485
a9 LANL.....IGY...A..DIWAKEREKLRNWIFTNCVKN..NYFIRYCGN.....TDDVDSSLLSAPLYGFI.EVSDSTFINTLTKIENDLK.....T.DVFV.KRYKT... 493
a10 LANE.....LGYKNEA..EEWINVAEEIKKDVFDNGYSEKLGYFTRYYG.....SNSVDSALLTLPLYGFI.DAKDPRFLRTFEKIERDLM.....ISDGLLL.RYKD... 501
a11 LAEA.....LGYKKEA..EEWKSVANEIKEDIMRHGIVE..GSFVRYYG.....GDEIDSALLTLPLYDFV.DANDKIFLNTLKRIESELK.....IEDGLYL.RYKK... 515
b8 LIES.....GD.ADGPLERWKQLRDDIHRDVCEKGYDKERNTFTQSYGS.....QAL.DASLLLIPQMGFL.PPDDKRVIGTIEAIQRELSTS.....DGFIL.RYPTDGE 522
b9 TLED.....RPELGGGDLDGWRALRDEVHREVCEKGYDPVRNTFTQYYGS.....REL.DASVLLIPRVGFL.PPDDPRVVGTVDAIGAELAR.....DGLV.RRYSTEGD 508
b10 AIEE.....FNLDGPIERWRELRAKLREEIMTNGFNEEIQSFTQCYDN.....TQV.DASLLQLAQIGFI.GFDDPKMLSTVARIEQELLDA.....HGFL.HRYHTDGS 526
b11 LAER.....QGEKSYA..QQWRAIADEIKADILEHG.VDSRGVFTQRYGD.....EAL.DASLLLVVLTRFL.PPDDPRVRNTVLAIADEL.TE.....DGLVL.RYR.VHE 542
b7 LAPR.....LGGHARP..QAWAAERDRIHATVLERGWSERRGAYTQAFDS.....DEL.DAAALLMPLYGFL.PATDARMNATIEAIARDL.TD.....DGLVL.RYRLIDD 515
a12 MARD.....LGLTA.PIKK.WKGIQDEIWNDVMTKGFDPQTNSFVQYYGS.....KNV.DASLLRLPILGFI.PANDERFIGTMKRIEDEL.....MVDGYLFRRYR...E 521
f8 IAFT.....RSLPC.PKHMVWMDTRDEIYEAIMDSGYNMEKGFFAQSFEN.....NDILDASVLVMPLVSFI.SPTDPRFLSTMDNIMKPLEKDG.....LMSNGLIF.RYN...N 552
b6 VLP.....GAAGRRCADLADTVLRETRRRCLHAD.GRWRRA.ED.....DDGPEAALL.VPMARGC.TFEDDGDAATRAYVESRLS.....EDGYLY.RFEHPGV 507
Family_2
a1 YDVG.....DEILSEDLSDTSGLLKGMVLKRMTYN.....GRFVRRIDEEN.....NQDLTVDSSLYA.PFFFGLVNANDKIMINTINEIESRL.....TVNGGII.RYENDMY 540
a2 RDIG.....DEVLASDMLDVAEYMKEEVLRRMVHN.....GRFIRRIDENG.....NKDLVIDASMYS.PYFFGMVDVRDPIMINTIRAIENSI.....KVSGGIA.RYENDMY 534
a3 NVFG.....DGDLSEKYLSAAERMYEAFEEKFYSEE.....YGYYARAIL.....DG.....KPDFTVDSALTSLVI.FGMKDPRDPKIVSTMEKISDTL.....WVNGVGGIA.RYQNDRY 569
a4 NVFG.....DPDLSEKYENAAERMYHAFDERFYSED.....TGYYARAII.....DG.....KPDFTVDSALTSLVL.FGMKDADDPKVISTMQRISEDL.....WVNGVGGIA.RYQNDRY 546
a5 RYFG.....EDELAQKYETVADQMRNSLD.LFWVGD.....HYARTIYMKEGQVH.....KIDETVDSSILLAPI.FNVIPMSDSRFVKNLETVIEKL.....SVKRGLV.RYEGDQY 527
a6 EFFG.....DENLKDKYLTAANEVKKGLE.RFYVGD.....HFARTIY....EDN.....SIDKTVDASTLFASI.LGPFDPKDPRVISNRKVVEEKL.....NINGGIA.RYENDWY 517
a7 KAVN.....KRDKVKDWGKTIEFLKHEV.....PKR.....FY.....LEDEERFA.....KSIN.PL...DKTIDTSILGLSYPFNLIDVDDERMIKTAEAIEKAF.....KYKVGGIG.RYPEDIY 523
Family_1
b1 DKLG.....EKRLAQNWKEAAGAIRRAVEEKCWDPERQ....VFYRGLKLMVPPEEWAHAEKEGKRGMIHVDEKGISRPMLWQDPVIDISLVGISVPFNLIAPDDERMVKMAEAIETHL.....TSPIVGGI.KRYENDPY 565
b2 DMEG.....DAAEATHLRQTADCWNEQIERWTFAGDQTMCAQLGISGHYVRIASPGTTDA.VQ.LSGMTPILNQTPDRAFLPTSDVISPDALAL.VRFGLRAPDDPHILATVKAIDHCLRAEL.....PQGPVWYRYTGDGY 590
b3 DACG.....KKDEATYLRETSDGWNDQLERWTYATDTAICRAAGVEGYYVRIAPPDSAEAGSP.KDGYVPIKNRPPGDTDRPAQEIVSPDALAL.VRFGLRSADDPRIVDTVKVIDAQLRCAL.....PQGPIWYRYNGDGY 590
b4 GKNG.....DAGSAAVYRATADEWQRSTEKWMFTTN....GPVGDGKYYLRISATGNPNDGAT.RD....WGN...GAGVHPENAVLDGGFLEF.VRLGVKAPADPYVADSLAETDASISQET.....PGGRMWHRYTYDGY 601
b5 EQNK.....DYESAQKYQEKADNWQKLIDNLTYTEN....GPLGNGQYYIRIAGLSDPDADF.....MINIAN...GGGVYDQKEIVDPSFLEL.VRLGVKSADDPKILNTLKVVDSTIKVDI.....PKGPSWYRYNHDGY 581
Sec_str hhhh.....hhhhhhhhhhhhhhhhhhhhh.....ssss.....ssss.....hhhhh-hhh.....hhhhhhhhhhhhhhssss.....ssss.....
of b5 .....aH7.....aS1.....aS2.....aH8.....aH9.....aS3.....aS4.....
Family_4
f5 KSFDDGDFASYLDSTASTLESYLSGSDGGFVNTDVNHIVENPDLLQQNSRQ.....GLDSATYIGPLLTHDIGESSST.PF...DVDNEYVLQSYYLLLEDNKDRYSVNSAYSAGAAI.....GRYPEDVY 378
f3 KALG..EECQAC.SVAPQILCHLQDFW....NGSAVLSNLPTNGRSGLDTN.....SLLGSIHTFDPAAACDDTTF...QPCSSRALSNHKLVVDSFRSVYGINNGRGAGKAA.....AVGPYAEDTY 338
f2 KTVG..SSCPYCDSQAPQVRCYLQSFWTGSYIQANFGGGRSG.....KDIN.....TVLGSIHTFDPQATCDDATF...QPCSARALANHKVVTDSFRSIYAINSGRAENQAV.....AVGRYPEDSY 336
f1 TAVG..SSCSWCDSQAPEILCYLQSFWTGSFILANFDSSRSG.....KDAN.....TLLGSIHTFDPEAACDDSTF...QPCSPRALANHKEVVDSFRSIYTLNDGLSDSEAV.....AVGRYPEDTY 311
Sec_str hhh.....hhhhhhhhhhhhhhhhh.....hhhhhhhh.....hhhhhhhhhhhhhh.....
of f1 .....H7.....H8.....H9.....


Family_6
a13 HYS.EYLYDATG..PSPPWIITTLFEAWYLEEIGKYN EVESILYWAYNHSQH.GLLPEAVNPSVNDSY.....PLPTTSPLTWSSAMFVI.V 589
Family_5
b12 ....DYDSNE.....WAFVDLRVATALKNMGKDK ESKVLIDWLTNQAHANFDLIPELLTRNNSDYF.....GSIPMAG...FGSGSY 646
Family_3
a8 ....DFMGEAKH.....PFLLTTLWLARVYVRLGRIE EAKNIISNIEEIAGNLKLIGEHADVEKKEFT.....GNFPQVFVHAEIINLV 560
a9 ....DFMGEAKH.....PFLLTTVWLARVYMRLGKID SAIEILNKINKVSRELHLVGEHVDVEKGEFT.....GNFPQIFVHAQLVIAI 568
a10 ....DFMGNVKH.....PFALVSTWLSRVYIRLGEVD RAKKVIEKLINCSTSSLLLAEHLDQNTCEPR.....GNFPQAFPHAGLIVAI 576
a11 ....DFLGSVVH.....PFALVTPWMARVYIRLNKVE DAIRLLEKLDKCSNSLKLLGEHIDQKNCEAR.....GNFPHSFPHAGIILSI 590
b8 AEGVDGLPGDEG.....AFLACSFWMADDLAMIGRVD EARKLFEKLLSLRNDLGLLAEEWDPRLQRQV.....GNFPQAFSHVPLIDTA 601
b9 P..VDGLPGGEG.....TFLVCSFWYADALHATGRTK EARELFERLVGLANDVGLLAEEYDPVAGHQL.....GNFPQAFSHVGLVNTA 585
b10 ....DGLAGDEY.....PFLICSFWLVEQYASSNRLD EAKEKMNRILAVQSPLGLLAEEYSTHHGRLA.....GNYPQAFSHIGLISAA 601
b11 .TD.DGLSGEEG.....TFTICSFWLVSALVEIGEVG RAKRLCERLLSFASPLLLYAEEIEPRSGRHL.....GNFPQAFTHLALINAV 619
b7 .RNLDGLEGGEG.....TFVLCSFWLAAALAQAGRPD RAQALFDRVAGFANDLGLLAEEIDPRTGELL.....GNFPQAFSHIGLINAA 593
a12 ...DDGLKGDEG.....SFLMLTFWYIEDLILMKRLK AARAALESIIEKANHLGLYSEEIDEKTGDFL.....GNFPQALSHLGIIRVA 597
f8 FVYEDGVGGDEG.....AFTMVCFWLVEALAMAGCSGYP KLLSTAVSMFEDLVRYSSHLQYYSEEC.SLSSESL.....GNSPQAFSSIAAIAAA 636
b6 P.....LGEAEG.....AFLLCGFAMALATHRVGDRV DAFRWFERTRAACGPAGLFAEEYDVRQRQLR.....GNLPQAFVHAMLLECS 581
Family_2
a1 .QRRKKQP....N....PWIITTLWLSEYYATINDKN KANEYIKWVINRALPTGFLPEQVDPETFEPT.....SVTPLVWSHAEFIIAI 615
a2 .KRVKSQP....N....PWIITTLWLAEYYLDLGQRE KALDYINWAMSRALPSGLLPEQVDPENFTST.....SVVPLVWSHAEFIITV 609
a3 .MRVKDDT....NIPGNPWIITTLWMARYYLRYGDFE RAWDLIMWVKSHRQKSGIFSEQINPYNGQPL.....SVSPLVWSHAEFIISL 648
a4 .MRVKDDP....SVPGNPWIITTLWMARYYMRFGDFE KAWNLIQWVKSHRQKSGIFSEQINPYNGEPL.....SVSPLVWSHSEFIISL 625
a5 .LRGGNNS....NI....WFISTLWLSQVYSLMGEKD KAKEKIDWVLSKSLPTGVIPEQIDDNDKYP.....SVSPLAWSHAELIRAI 601
a6 .LKQDEKS....NA....WFITTLWLAQQYILEGNKE KAKKYIDWVISHMLPTGIIPEQVSPKNTYP.....SVAPLVWSHAEFVKTI 591
a7 .............FGGNPWIITTLWLSLYYRRLYKVLKEKDDNGA DIYLQKSKKLFNWVMKYSFDGLFPEQIHKELGVPM.....SAMPLGWSNAMFLIYV 606
Family_1
b1 .............VGGNPWILTTLWLAQLYIKQGR IKKALNHLKWVVDHRTDLDLLPEQVDRETGKAA.....WVVPLTWSHAMFVLTV 636
b2 GEHADGTAFDGTGQ.GRPWPLLTGERAHYELAAGRRGAAE ALLTTMEKSAGQSGLFPEQVWDQPDLPDRELFYGSPSGSAMPLVWAHAEHIKLL 683
b3 GEHQDGTPFDGTGQ.GRPWPLLAGERAHYELAAGRKDEAA NLLKALEDSAEPGGLLPEQVWDGADLPERQLLHGRPSGSAMPLVWAHSEHIKLL 683
b4 GEKADGSPWDGTGI.GRLWPLLSGERGEYALANGQ..DAL PYLETMHSAANAGYMIPEQVWDRDEPTSYGHELGRSTGSASPLSWAMAQYVRLA 692
b5 GEPSKTELYHGAGK.GRLWPLLTGERGMYEIAAGK..DAT PYVKAMEKFANEGGIISEQVWEDTGLPTD.....SASPLNWAHAEYVILF 663
Sec_str ....hhhhhhhhhhhhh.....hhhhhhhhh.....hhhhhhhhhhhhh
of b5 ....aH10.....aH11.....aH12..
Family_4
f5 NGDGSSE........GNPWFLATAYAAQVPYKLVYD.AKSASNDITINKINYDFFNKYIVDLSTINSGYQSSDSVTIKSGSDEFNTVADNLVTFGDSFLQVILDHINDDGSLNEQLNR.....YTGYST.SAYSLTWSSGALLEAI 509
f3 QG.............GNPWYLTTLVAAELLYDALYQWDKQGQVNVTETSL..PFFKDLSSNVTT..GSYAKSSSA.....YESLTSAVKTYADGFISVVQEYTPDGGALAEQYSR.....DQGTPV.SASDLTWSYAAFLSAV 453
f2 YN.............GNPWFLTTLAAAEQLYDALYQWDKIGSLAITDVSL..PFFKALYSSAAT..GTYASSTTV.....YKDIVSAVKAYADGYVQIVQTYAASTGSMAEQYTK.....TDGSQT.SARDLTWSYAALLTAN 451
f1 YN.............GNPWFLCTLAAAEQLYDALYQWDKQGSLEVTDVSL..DFFKALYSDAAT..GTYSSSSST.....YSSIVDAVKTFADGFVSIVETHAASNGSMSEQYDK.....SDGEQL.SARDLTWSYAALLTAN 426
Sec_str ....hhhhhhhhhhhhhhhhhhh.ssss.....hhhhh.....ssss...hh-------hhhhhhhhhhhhhhhhhhhh.....hhhhhhhhhh
of f1 ....H10.....S1.....H10'.....S2.....H11.....H12....

```
Family_6
      a13 SLHYKQEHSSVSILIFLVIVIVLIPILYYLIKKFYHS.............   626
Family_5
      b12 ILSINNYYNNLR.....................................   658
Family_3
       a8 KELEG...NKNNVT...................................   571
       a9 KELNDTLTDKNII....................................   581
      a10 VELEERLV.........................................   584
      a11 IELEEKLNAENITTNK.................................   606
       b8 LRL....TASGAYGG..................................   612
       b9 LALFGGEEAG.......................................   595
      b10 RAINFEEARNR......................................   612
      b11 VHVIRAEEEADSSGMFQPANAPM..........................   642
       b7 AAIDQANGALPRHDA.DPAH.............................   612
      a12 PKLEEAFLKRVSKINE.................................   613
       f8 HILDKA.LSIFNEI...................................   649
       b6 VRLAGESTLP.......................................   591
Family_2
       a1 NKLLNHI..........................................   622
       a2 DKLISR...........................................   615
       a3 LEYNEYVRNHS......................................   659
       a4 LEYSDLIRNRS......................................   636
       a5 ..Y..ALRN.GTLDPK.................................   612
       a6 ..Y..YFKQLG......................................   598
       a7 YENDKV..IIP......................................   615
Family_1
       b1 LDLLEIGAIGPKKS...................................   650
       b2 RSLADGSVFDMPPQGVERYIRNKTPSHLRIWCFNNTISAMPTGKVLRLEL   733
       b3 RSLRDGAVFDMPPQGVKRYIEAKTVSPFRTWRFNNKIRALPEGKTLRIEL   733
       b4 AGVKAGAPVETPQNVAARYAAGTPLSSPELSVTAPEALSTADSATAVVRG   742
       b5 ASNIEHKVLDMPDIVYKRYVAK............................   685
Sec_str   hhhhhh......hhhhhhhhhh
of b5     ..............aH13
Family_4
       f5 RLRNKVKALA........................................   519
       f3 GRRNGTVPASWGSSTANAVPSQCSGGTVSGSYTTPTVGSW..........   493
       f2 NRRNAVVPAPWGETAATSIPSACSTTSASGTYSSVVITSW..........   491
       f1 NRRNSVVPASWGETSASSVPGTCAATSAIGTYSSVTVTSWPSIVATGGTT   476
Sec_str   hhh.........hhhh
of f1     ............H13
```